

**Automatic Guitar Tablature Transcription
from Audio using Inharmonicity Regressions
and Bayesian Classification**

Jonathan Michelson

July 2017

School of Music
Carnegie Mellon University
Pittsburgh, PA 15213

Thesis Committee:

Dr. Richard Stern, Dr. Thomas Sullivan

*Submitted in partial fulfillment of the requirements
for the degree of Master of Science in Music and Technology*

Abstract

We propose and evaluate two new methods to classify guitar strings with the goal of developing automated tablature transcription. The classifiers use only audio input and are based on the measured inharmonicity of the sounds produced by the strings. While previous inharmonicity-based systems were able to obtain good performance classifying guitar strings based on sample data produced by individual guitars, we attempt to develop a system that can obtain improved string-identification performance for guitars on which the system was not individually trained, using training data aggregated from multiple guitars of the same type. We also derive and evaluate an inharmonicity compensation feature to allow for tablature transcription of arbitrarily-tuned guitars, provided the tuning is known.

The first method linearly regresses the log-inharmonics of guitar strings against their pitches to obtain characteristic trajectories, then assigns unseen notes to the strings whose means and variances of their trajectories maximize the probability of the measured inharmonics of the notes. The second method, developed as a baseline against which to compare the first and which we refer to as simply Bayesian classification, characterizes the inharmonicity distribution of each fretboard position as a normal probability density, then similarly assigns unseen notes to the fretboard positions that maximize the likelihood of their observed inharmonics. Results from classical and acoustic guitars in the standard Real World Corpus of guitar recordings show that our regression method generally performs better than the existing inharmonicity-based system for test recordings on which the system was not individually trained, as well as show that leveraging the linear structure of log-inharmonicity trajectories increases overall performance compared to the baseline method. Also, results from personally-recorded acoustic and electric guitars in various alternate tunings show that our tuning compensation is successful at predicting tuning-related inharmonicity changes and improving transcription performance.

Acknowledgments

My sincere gratitude goes to Rich Stern for advising me every step of this journey. I'd never before tackled a research project of this scope, but Rich's guidance, patience, and insight somehow made this thesis happen. It's been an immeasurably rewarding experience, and I owe that to him. I'll miss our weekly one-on-one meetings that evolved into friendly chats about music, family, and life.

I'm indebted to Tom Sullivan for lending his qualified examination of this thesis. I'm also indebted to him in the literal sense as he welcomed me to Pittsburgh by treating me to lunch. His hospitality and accessibility helped me know I was in the right place when I arrived two years ago.

I'd like to thank Matthieu Hodgkinson for sharing with me his MAT algorithm code, as well as Jakob Abeßer and Isabel Barbancho for their helpful correspondence.

My peers Che-yuan Liang and Raymond Xia have both, of their own accord, contributed meaningfully to this work through our many brainstorming and discussion sessions, and for them I'm grateful. Many thanks to Nick Pourazima, who lent his Fender Telecaster for some of our experiments.

I need to thank the remainder of my CMU graduate music technology cohort – Devansh Zurale, Steven Krenn, Chen Liang, Mutian Fu, Hao Huang, Anirudh Mani, Garrett Osborn, Nick Pourazima, Mingyuan Yu, and Yixuan Zhang – for the support role they played by simply being the fun constructive community that they were.

Finally, I acknowledge my mom and dad, to whom all success is due. I'd also like to thank my brother Nick Michelson solely for persuading me to abandon writing a grandiose Acknowledgments section, and definitely nothing else.

Contents

1	Introduction	1
2	Previous Work	5
2.1	Automatic Tablature Transcription	5
2.1.1	Graph Search Methods	5
2.1.2	Supervised Learning Solutions	6
2.1.3	Multimedia Approaches	6
2.1.4	The Partial Coincidence Tally (PCT) Method	6
2.2	Inharmonicity Motivation	7
2.3	Inharmonicity Estimation	8
3	Proposed Transcription Methods	11
3.1	Inharmonicity Trajectories	11
3.1.1	Motivation	11
3.1.2	Inharmonicity Estimation	14
3.1.3	Log-inharmonicity	15
3.1.4	Learning Inharmonicity Trajectories	19
3.1.5	String Classification	21
3.1.6	Tablature Conversion and Refinement	23
3.2	Baseline Bayesian Classifier	24
3.2.1	Motivation	24
3.2.2	Inharmonicity Estimation	24
3.2.3	Fretboard-Position Inharmonicity Distributions	25
3.2.4	Fretboard Position Classification	25
3.3	Tuning Compensation Feature	26
3.4	System Overviews	28
4	Experiments and Results	33
4.1	RWC Evaluation	33
4.2	Tuning Compensation	40
5	Discussion	43
5.1	Analysis	43
5.2	Limitations	60

5.3	Future Work	61
6	Summary	65

List of Figures

1.1	Musical score and two identical realizations of that score in tablature notation. Figure taken from [?].	2
2.1	Power spectrum of note D2 plucked on an electric guitar at string 5 and fret 5. Red dashed lines are drawn at integer multiples of the fundamental. The partial peaks begin visibly skewing rightward after the fifth harmonic.	8
3.1	Example inharmonicity trajectories of acoustic guitar strings, calculated from measured open-string inharmonicities. Only Fret 0 (open-string) through Fret 3 are shown. Observe the restriction to the exponential trajectory defined by equation (3.1).	12
3.2	Median-adjustive-trajectory (MAT) method flowchart.	16
3.3	Inharmonicity estimates for Frets 0-12 on Strings 1-6 of RWC Acoustic Guitar AG111. Note the exponential trajectory as in Fig. 3.1 and the increased segregation since notes and their inharmonicities are now plotted against their fundamental instead of their fret number.	17
3.4	Log-inharmonicity estimates for same frets, strings, and guitar as Fig. 3.3.	18
3.5	Comparison of string inharmonicity characterization approaches for our recorded Fender Telecaster. Top: <i>theoretical</i> trajectories. Note how each line begins precisely at the first marker of each string (i.e. the open-string note) and continues along a trajectory that ignores empirical inharmonicity estimates. String 1 most clearly shows this. Additionally, the trajectories for String 3 and String 4 are effectively co-linear and therefore indiscriminable. Bottom: regressions, or <i>empirical</i> trajectories. Each regression was obtained by minimizing the bisquare weighted error of the residuals of the string, so the lines more closely follow the empirical inharmonicity scatterplot estimates.	20
3.6	Effect of alternate tuning on log-inharmonicity trajectories of an electric guitar. The circles are log-inharmonicity estimates of Frets 0-12 on String 6 of RWC Electric Guitar EG131 in standard tuning, and the dashed line is their bisquare-weighted regression. The asterisks are log-inharmonicity estimates for the same frets, string, and guitar recorded two semitones down. The dotted line passing through them is the <i>predicted</i> regression obtained by applying equation (3.25) to the dashed line.	29

3.7	System overview of tablature transcription with log-inharmonicity trajectories using probabilistic classification.	30
3.8	System overview of tablature transcription with Bayesian fretboard-position classification.	31
5.1	Log-inharmonicity regressions learned from the 12 recordings of each acoustic and electric guitar and the 6 selected recordings of each electric guitar superimposed over their inharmonicity measurements.	46
5.2	Superimposed string-wise regressions of all three RWC classical guitars. Top: Strings 1, 2, and 3. Bottom: Strings 4, 5, and 6. Plots were separated for visibility, and axes scales were kept identical for comparison. Observe the consistent intercept and slope despite varying guitar (with the only exception being String 6).	47
5.3	Superimposed string-wise regressions of all three RWC acoustic guitars. Top: Strings 1 and 2. Bottom: Strings 3, 4, 5, and 6. Plots were separated for visibility, and axes scales were kept identical for comparison. Observe the consistent intercept and slope despite varying guitar.	48
5.4	Superimposed string-wise regressions of all three RWC electric guitars. Top: Strings 1, 2, and 3. Bottom: Strings 4, 5, and 6. Plots were separated for visibility, and axes scales were kept identical for comparison. Observe the <i>inconsistent</i> intercept and slope. In particular, note the overlap between one of the regressions corresponding to String 3 (dashed red) and the regressions corresponding to String 2. Also observe the overlapping slope and intercept ranges for Strings 3 and 4 across the two plots.	49
5.5	box plots of log-inharmonicity estimates for 25 uniform plucks of the six open-string notes on our acoustic guitar (Takamine G-series). Horizontal red lines are medians, horizontal blue lines are upper and lower quartiles, whiskers denote data range excluding outliers, red “+”s are outliers.	51
5.6	box plots of log-inharmonicity estimates for 25 uniform plucks of the six open-string notes on our electric guitar (Ibanez RG). Horizontal red lines are medians, horizontal blue lines are upper and lower quartiles, whiskers denote data range excluding outliers, red “+”s are outliers.	52
5.7	Best and worst bisquare log-inharmonicity regression fits, as measured by r^2 , for representative guitars from each of the three classes. Top row: classical. Middle row: acoustic. Bottom row: electric. Left and right columns are best and worst fits, respectively. Dashed red lines, only visible on the bottom-right plot, denote 95% confidence intervals.	53
5.8	box plots of log-inharmonicity regression residuals for EG131 by dynamic level. Group 1: piano, Group 2: mezzo, Group 3: forte. Red center horizontal lines are medians; top and bottom edges of blue boxes represent upper and lower quartiles; whiskers depict full range of data excluding outliers; red plus signs denote outliers. Vertical axes not constant throughout plots; we’re highlighting relative distributions of the box plots here.	55

5.9	Effects on log-inharmonicity measurements of various electric guitar geometry parameters. In all box plots, measurements are from 25 uniformly-played open notes (G3) on string 3 in standard tuning. Horizontal red lines are medians, horizontal blue lines are upper and lower quartiles, whiskers denote data range excluding outliers, red “+”s are outliers. Note the changes in scale on the vertical axes.	58
5.10	box plots of log-inharmonicity measurements of 25 open-note plucks on a new, 0.010-in gauge, unwound third string, processed with two levels of drive in Logic Pro’s stock overdrive plug-in.	60
5.11	Solid black lines and left vertical axis: probability of correctly transcribing a note given the residual width under which the log-inharmonicity of the note falls. Dashed orange lines and right vertical axis: CDF of the log-inharmonicity residuals, so one can get a sense of the proportion of data represented by the horizontal axis.	64

List of Tables

4.1	Error probability comparisons between the baseline classifier (Bayes), the existing partial coincidence tally method (PCT), and our proposed probabilistic regression classifier (Regr.). PCT results are taken from [?]; we’ve excluded error due to f_0 estimation in their results. Note that PCT results for electric guitars are intentionally excluded to avoid potentially unfair comparison, while PCT results for the “Guitar-independent” scenario are not available.	35
4.2	Error probability comparisons. Results from our baseline (“Bayes”) and regression (“Regr.”) methods which differ from PCT to a statistically-significant ($\alpha = 0.05$) degree are shaded in gray. Electric guitar trials and “Guitar-independent” scenarios are omitted since comparisons in those cases are not available.	37
4.3	Error probability comparisons. Superior results in trials with statistically-significant ($\alpha = 0.05$) performance differences are shaded in gray. PCT results omitted to highlight performance differences between our baseline and regression methods.	38
4.4	String-wise F1-scores for the RWC classical guitars, using regression classification for the “Guitar-averaged” scenario.	38
4.5	String-wise F1-scores for the RWC acoustic guitars, using regression classification for the “Guitar-averaged” scenario.	39
4.6	String-wise F1-scores for the RWC electric guitars, using regression classification for the “Guitar-averaged” scenario.	39
4.7	String-wise F1-scores for our Fender Telecaster at various tunings. Regressions trained on recordings of all three RWC electric guitars. WSU: whole-step up (from standard); WSD: whole-step down (from standard); orig.: no tuning compensation used in this trial; comp.: tuning compensation used in this trial. The “Overall” row displays average F1-score of each column. Statistically significant ($\alpha = 0.05$) results highlighted in gray.	41
4.8	String-wise F1-scores for our Takamine G-series at various tunings. Regressions trained on all recordings of all three RWC acoustic guitars. WSU: whole-step up (from standard); WSD: whole-step down (from standard); orig.: no tuning compensation used in this trial; comp.: tuning compensation used in this trial. The “Overall” row displays average F1-score of each column. Statistically significant ($\alpha = 0.05$) results highlighted in gray.	42

5.1	Confusion matrix for one classification trial with EG131. Strings 3 and 4, as they are here, were frequently misclassified because of the poor discriminability of their corresponding regressions.	50
-----	---	----

Chapter 1: Introduction

The guitar is a popular musical instrument whose family comprises a diverse collection of stringed instruments. Its most prominent classes are the acoustic, classical, and electric guitars, and intraclass variation is high in characteristics such as body shape, string material, and amplification modality. Despite this diversity, the majority of guitar configurations exhibit some unifying features, like having six strings tuned to $E2$ (82 Hz), $A2$ (110 Hz), $D3$ (147 Hz), $G3$ (196 Hz), $B3$ (247 Hz), and $E4$ (330 Hz), for example. Performance is relatively uniform as well; to play, a performer presses these strings with one hand against the metal frets of the neck, each of which increase the pitch of the string by one semitone, while strumming or plucking them with the other hand. Guitars typically have upwards of 20 frets, allowing for versatile musical passage realizations along the fretboard.

A consequence of the open-tunings and extensive fretting locations of the strings is liberal pitch overlap between neighboring strings; multiple fretboard locations can usually be selected to produce a given pitch. Conventional music scores that represent passages as notes and chords therefore fail to communicate fretboard position. This ambiguity renders scores unfit for guitar students trying to learn more skillful placement of scales and arpeggios, or for enthusiasts trying to uncover the fretboard positions used on riffs recorded by virtuosic players. Figure 1.1 illustrates this ambiguity.

Tablature, on the other hand, is an alternative music notation that does not suffer from the one-to-many mapping of scores. The staff, instead of representing pitch as in classical scores, depicts a birds-eye view of the guitar neck from the perspective of the performer. Each of the six

horizontal lines signify a corresponding string on the guitar, and numbers on each line specify the fret to be fingered. Time progresses from left to right as in scores.

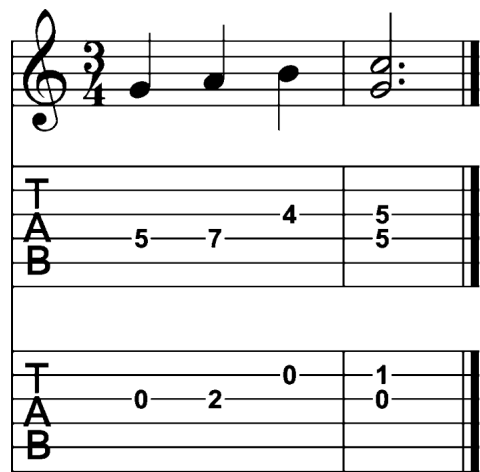


Figure 1.1: Musical score and two identical realizations of that score in tablature notation. Figure taken from [?].

Tablature (abbr. tab) is more popular than scores for a number of likely reasons [?]. Its accessibility is attractive to those without formal musical training; interpreting a pictorial representation of the fretboard poses a smaller learning curve than studying music theory and deciphering score notation. Additionally, the lines and numbers in tablature that represent the strings and frets are easily represented with ASCII characters, rendering it an exceptionally portable mode of communication. Indeed, one can create, modify, and transfer tablature files using simply a text editor.

Accurate tablature transcription is a laborious task, however. Dependable tabs containing few errors typically require the careful listening of a seasoned musician to an audio recording, and possibly video recordings for additional consultation. Reliable automation of this transcription task would benefit the guitar student community by expediting quality tab generation and unearthing fretboard positions of guitar riffs in songs for which accurate tablature does not yet exist. Recent technologies proposed to encourage proliferation of accessible tab transcription systems, such as web framework Robotaba [?], demonstrate a desire among MIR researchers to see this theoretically mature but practically nascent field mature into student-benefitting applica-

tions.

An important feature in systems that transcribe tablature from audio input is inharmonicity. Briefly, inharmonicity is a phenomenon that skews the harmonic partials of a vibrating string upwards in frequency and whose variation trend is somewhat unique per string. Because of this, it is highly discriminative and has successfully been used as the sole feature in a transcription system [?] which we refer to as the existing inharmonicity-based method. This is discussed further in Chapter 2, but for now is introduced simply for context.

This work builds on the existing method. We propose an alternative procedure to exploit the inharmonicity of an unseen note for string classification: we assign to it the string whose pre-computed log-inharmonicity trajectory most probably predicts the estimated log-inharmonicity of the note. To contextualize its performance as well as that of its predecessor, we also introduce a baseline Bayesian classifier of fretboard position given an unseen note. The measured inharmonicities of each fretboard position are collected and fit to a normal distribution, and are used to return the string that maximizes the probability of having observed the measured inharmonicity of the unseen note. To the author’s knowledge, no results for strictly inharmonicity-based systems have yet been reported from a simple classifier for the sake of benchmarking.

We evaluate the performances of these two systems in three scenarios, comparing their performance to that of the existing method in [?]: when the systems have been trained on recordings of only the test guitar, when the systems have been trained on recordings of multiple guitars including the test guitar (all of which were in the same classical, acoustic, or electric class), and when the systems have been trained and tested on separate sets of guitars. Results show that although the existing system excels in the first case, our regression method surpasses it in the second case on average, suggesting its greater generalizability. The results also show that our leveraging of the linear log-inharmonicity string trajectories yields performance gains over the baseline fretboard-position classifier.

Additionally, this work introduces an inharmonicity compensation factor that allows for successful transcription of guitars whose tunings differ from those of the training set, provided that

the alternate-tunings are known. Experiments on a personally recorded Fender Telecaster and a Takamine G-series in various tunings reveal that our compensation factor can restore the transcription performance of an arbitrarily-tuned guitar to that of its more accurately transcribed standard-tuned trial.

In the next chapter, we survey previous work in the field of automatic tablature transcription, highlighting the existing inharmonicity-based method. This is followed by a more substantial motivation of the suitability of inharmonicity for the string discrimination task and a review of the literature on inharmonicity estimation. In Chapter 3, we introduce both systems in detail. Chapter 4 reports the results of our experiments, and Chapter 5 contains analysis of the results, limitations of significance, and future work. We summarize this thesis in Chapter 6.

Chapter 2: Previous Work

Music transcription is a rich research topic with many contributors. We begin with a brief review of various teams who have focused specifically on tablature transcription. We then introduce a feature of guitar audio referred to as inharmonicity, which has been used exclusively and successfully in one system, whose detailed summary we defer to Chapter 3. We close this chapter with a survey of the inharmonicity estimation literature.

2.1 Automatic Tablature Transcription

2.1.1 Graph Search Methods

A popular framework for the stringed-instrument transcription task, introduced by Sayegh [?] for both transcription and fingering problems, is that of weighted directed graphs that capture all plausible sequences of fretboard positions for a given musical passage. Edge weights in the graph are typically informed by mechanical or musical desirability of the fretboard sequences implied by the node transitions. Radicioni [?] used this graph paradigm to obtain optimal fingerings from music scores. Yazawa [?] extracted pitch and melody information from guitar recordings, and computed an optimal fingering based on plausible multipitch estimation results. Burlet [?] developed a polyphonic extension to monophonic optimal search algorithms. Burlet [?] also introduced a guitar-specific deep belief network whose polyphonic transcription output was arranged into tablature with a graph search routine. Radisavljevic [?] proposed a method to learn the optimal weights on the graph given optimal tablature.

2.1.2 Supervised Learning Solutions

A number of teams have used machine learning techniques to ascertain fretboard position or related parameters. Gagnon [?] proposed using neural networks on scores to deduce general fretboard region position and number of strings playing, while Tuohy [?] used a genetic algorithm on scores to obtain tablature. Barbancho [?] attempted guitar string classification with a Fisher linear discriminant using a multitude of time and frequency-domain features. Barbancho [?] used HMMs on polyphonic recordings of common guitar chord sequences to transcribe their fret position and fingering. Abesser [?] and Kehling [?] achieve good string transcription performance after training an SVM on standard-tuning guitar audio and using a slew of partials-related features, while Dittmar [?] adapted [?] into a real-time implementation.

2.1.3 Multimedia Approaches

Some researchers have explored usage of media other than audio or musical information (i.e., scores). O’grady [?] exploited hexaphonic pickups and matrix factorization of their output to record fretboard positions. Paleari [?], Burns [?], and Kerdvibulvich [?] extract tablature from video recordings using image processing techniques.

2.1.4 The Partial Coincidence Tally (PCT) Method

One particular system inspired this work. Barbancho [?] realized the discriminative potential of a singular feature known as inharmonicity, and devised an algorithm that exclusively used inharmonicity to achieve impressive transcription results using only audio as input. We discuss inharmonicity in detail and methods for its estimation in the remainder of this chapter. Their method, which we refer to as the partial coincidence tally (PCT) method, is also given a more thorough treatment later in Chapter 3.

We were particularly interested in the PCT method for a number of reasons. Firstly, its input

is audio. Among the most interesting use cases for an automatic tab transcription system is direct audio-to-tablature capability, as audio is the natural output of the object we're interested in transcribing. Other inputs, like scores or video, are to an extent already a form of transcription that detracts from the impressiveness of a tablature transcriber. Secondly, because PCT examines the underlying physical signatures of notes, namely their inharmonicities, it can theoretically decode the fretboard position of any note despite how unexpected its location may be. Other methods, like graph searches, rely on optimization of cost functions that are generalizations of tab playability rules, and therefore are not as theoretically well-suited to correctly transcribe aberrant note sequences. Lastly, their method exclusively uses inharmonicity. Other systems that use inharmonicity [10, 11, 12, 13] exploit additional features that aid in string discriminability. We were captivated by the idea of singularly using inharmonicity to achieve impressive transcription accuracies.

2.2 Inharmonicity Motivation

Inharmonicity is key in many string discrimination systems [10, 11, 12, 13]. When an ideal string fixed at both ends is displaced, the restoring force that causes it to oscillate is its tension. However, for a real string with stiffness, its elasticity contributes to this restoring force [14], causing the frequencies of the modes of vibration to deviate from integer multiples of the fundamental. Instead, the solutions to the one-dimensional wave equation become:

$$f_k = kf_0\sqrt{1 + \beta k^2} \quad (2.1)$$

where f_k is the k th harmonic of fundamental f_0 and β is the inharmonicity of the string, defined by

$$\beta = \frac{\pi^3 Q d^4}{64 T l^2}. \quad (2.2)$$

In words, the inharmonicity β of a vibrating string is a dimensionless quantity that depends on the Young’s modulus Q , diameter d , tension T , and vibrating length l of the string, and which skews the k th partial of the string upwards in frequency according to equation (2.1). See Figure 2.1. Note that for an ideally harmonic string, $\beta = 0$ and (2.1) reduces to $f_k = k f_0$, aligning with the ideal integer-multiple relationship between a fundamental and the locations of its harmonics.

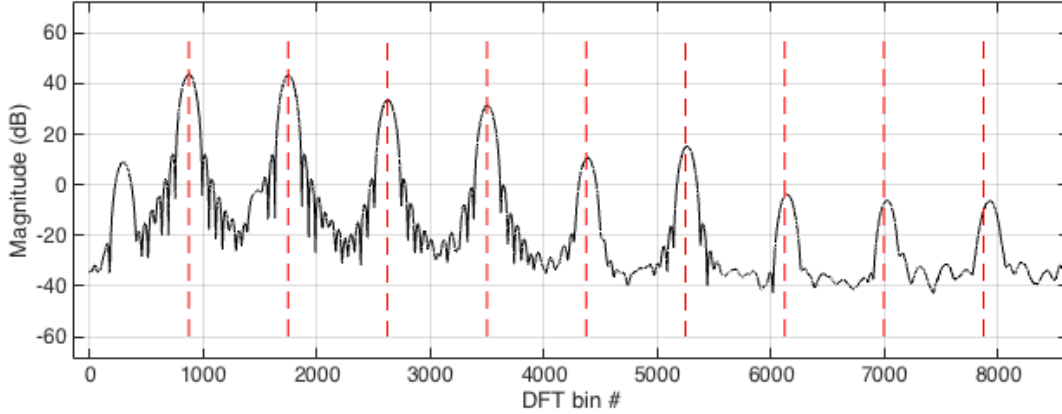


Figure 2.1: Power spectrum of note D2 plucked on an electric guitar at string 5 and fret 5. Red dashed lines are drawn at integer multiples of the fundamental. The partial peaks begin visibly skewing rightward after the fifth harmonic.

The suitability of inharmonicity for the string discrimination task can be understood through equation (2.2). The six strings of any guitar necessarily exhibit different combinations of Q , d , and T due to varying thickness, material, and tuning. Each string therefore has an “inharmonic signature” which presumably distinguishes it from the others. Capturing the pattern of the inharmonicity variation of a string along its frets should thus reveal sufficiently identifying information.

2.3 Inharmonicity Estimation

Empirical estimation of inharmonicity is itself a topic of research which merits a brief review. Using pitch extraction techniques to perform estimation was first attempted by Galembo in 1979 and 1986 [? ?]. Several additional methods have been proposed since then. In 1994, Galembo [?

] hypothesized a connection between partials-based fundamental pitch estimates and the degree of inharmonicity in a spectrum. They specifically investigated cepstral analysis and a variant of the harmonic product spectrum method applied to both synthetic tones and recorded piano notes. They found that, because these two techniques leverage periodicity in the frequency domain to produce their pitch estimates, inharmonicity influenced the quality of their estimates in a deterministic manner: more inharmonicity produced wider and less focused pitch estimate lobes, and from quantifying this relation an inharmonicity estimate could be obtained.

The same authors in [?] introduced another method in 1999, dubbed the inharmonic comb filter (ICF) method. To estimate the inharmonicity of a note, they first applied numerous comb filters to its spectrum. The locations of the notches of the comb filters were parametrized with a range of inharmonicity values that sufficiently sampled the expected inharmonicity range of the note. The ICF whose inharmonicity best matched that of the note would produce the output with lowest spectral energy, since its notches would most align with, and therefore best cancel, the inharmonic note partials. The inharmonicity parameter associated with this output-minimizing ICF was thus selected as the estimate.

Rauhala [?] introduced an efficient procedure in 2007 that iteratively catalogued estimates of the deviations of partial frequencies of a note and returned increasingly better estimates of the inharmonicity. The algorithm, known as the partial frequency deviations (PDF) method, is initialized with a reasonable first estimate of inharmonicity, and then the spectrum of the note is searched for partial peaks. Differences are measured between the locations of these discovered peaks and those of the expected peaks, and the aggregate deviation trend is used to refine the inharmonicity estimate – a majority of positive differences implies the inharmonicity should be increased, and a majority of negative differences implies the opposite.

In 2009, Hodgkinson [?] proposed a yet more efficient and accurate algorithm, named median-adjustive trajectories (MAT). Their routine exploits the fact that inharmonicity can be estimated with any two partials and their corresponding indices in the harmonic series. Pairs of low-index partials are first considered, since they are minimally affected by inharmonic skew

and therefore have high location reliability. With these initial inharmonicity estimates, additional partials are collected and used to refine the previous inharmonicity estimates, with which more partials are located, so on and so forth. MAT was the inharmonicity estimation routine employed in this work, and is discussed at length in Chapter 3.

Three years later, Barbancho [?] introduced another PFD-based approach as part of a guitar tablature transcription system. They started by cataloguing the deviations of the first 10 partials, then fitting a polynomial to the PFD curve. They derived the relation between the coefficients of the polynomial and the inharmonicity, and returned an initial inharmonicity estimate according to this derivation. Then using this inharmonicity to aid in their search for higher-index partials, they catalogued a larger number of deviations of partials than that of the first iteration, and performed the same polynomial fit and inharmonicity derivation routine. This would refine the inharmonicity estimate, allowing for increasing numbers of partials for which to search.

Also in 2012, Abesser [?] used parametric spectral modeling, again in the context of a guitar tablature transcription system. To estimate the inharmonicity of a plucked guitar note, they modeled each audio frame with the autoregressive (AR) filter that best approximated its spectral content. They obtained the locations of poles and the locations of ideal harmonics, and like Barbancho [?] fit the PFD curve with a polynomial from whose coefficients an inharmonicity estimate could be derived.

The MAT algorithm was chosen for its efficiency and accuracy over the algorithms that preceded it. In [?], benchmark tests are run against the PFD method of [?] (which itself is more efficient than the earlier ICF method of [?]), and results shows that MAT is about an order of magnitude quicker and more precise on the test set of synthesized inharmonic tones. To our knowledge, no comparisons exist between MAT and the two methods developed after it as part of tablature transcription systems (modified PFT [?] and AR filter modeling [?]). Though since we were having trouble replicating the two more recent systems and the author of [?] was willing to share the algorithm code, we decided to use MAT for this work.

Chapter 3: Proposed Transcription Methods

In this chapter, we present both our inharmonicity trajectory method and our baseline method in detail. More time was spent innovating and developing the former, so we appropriately give it a lengthier treatment. We begin with the motivation for the trajectory method, then we present its various components: the specific inharmonicity estimation algorithm we used, transformation of inharmonicity to log-inharmonicity, learning log-inharmonicity trajectories, string classification using the learned trajectories, and the final tablature conversion and refinement process. Next, we discuss the motivation for the baseline classifier, its inharmonicity estimation algorithm, and its fretboard-position classification mechanism. We finally introduce our feature that allows compensation for arbitrary tuning before concluding with overviews of the two presented systems.

3.1 Inharmonicity Trajectories

3.1.1 Motivation

Barbancho [?] derived that inharmonicity varies in a deterministic manner with respect to fret positions along a given string. Specifically, they showed that the inharmonicity $\beta(s, n)$ of a note produced by string s at fret number n could be expressed in terms of the inharmonicity of the open-string note $\beta(s, 0)$ according to:

$$\beta(s, n) = \beta(s, 0)2^{\frac{n}{6}} \quad (3.1)$$

In other words, the variation in inharmonicity along a given string can be expressed simply in terms of its open-note inharmonicity, effectively defining a restricted trajectory along which the inharmonicity must vary. Figure 3.1 illustrates these trajectories.

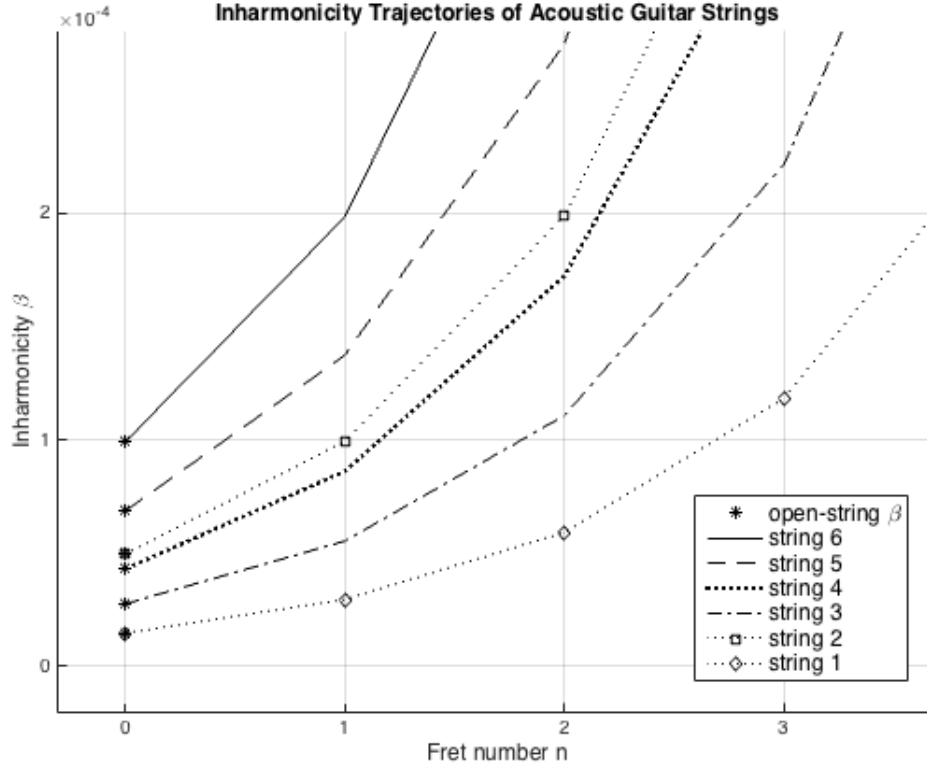


Figure 3.1: Example inharmonicity trajectories of acoustic guitar strings, calculated from measured open-string inharmonicities. Only Fret 0 (open-string) through Fret 3 are shown. Observe the restriction to the exponential trajectory defined by equation (3.1).

Recognizing this, they were able to successfully transcribe string and fret number of an unknown guitar note. The essential mechanism of their system was a similarity-determination routine $A(X, \hat{\beta}_s)$ that quantified the resemblance between the spectrum X of an unknown note and the expected inharmonicities $\hat{\beta}_s$ of candidate strings s from which the note could have originated. First, to estimate the inharmonicity of an open-note, they followed the procedure described in section 2.3 in which they estimated its fundamental, then catalogued the deviations of its partials from their ideal harmonic locations, and fit to them a polynomial from whose coefficients

they derived the inharmonicity estimate. Next, they inferred the corresponding fret on which the note could be played for each candidate string, and used these string-fret candidate numbers in equation (3.1) to obtain expected inharmonicity values of the unknown note for each candidate string. In other words, they followed the theoretical inharmonicity trajectory of each candidate string to arrive at an expected inharmonicity value at the appropriate fret. Lastly, they applied their similarity determination algorithm $A(X, \hat{\beta}_s)$, which went as follows. Given the expected inharmonicity of each candidate fretboard position, they obtained the corresponding expected skewed frequency locations at which to find the inharmonic peaks, and searched the spectrum of the unseen note for them using search windows centered at those expected locations. They performed this search for a set number of peaks, and catalogued their deviations from the search window centers. Their motivation was that the expected inharmonicity of the true fretboard position would produce a spectrum with inharmonic peaks that were measurably more coincident with the inharmonic peaks of the unknown note than were any of the imposter strings. They found that, indeed, the theoretical inharmonicity of the fretboard position on which the note was truly plucked would reliably yield the largest number of coincident peaks located inside the spectral search window bounds, hence our name “partials coincidence tally” for their algorithm. Their work demonstrated the sufficiency of inharmonicity as a discriminating feature.

This method capitalizes further on their trajectory realization. Rather than zooming into the spectra of fretboard positions to evaluate their candidacy, we propose characterizing each candidate string by its inharmonicity trajectory and classifying a note as simply belonging to the string whose inharmonicity trajectory evaluated at the fundamental of the unseen note best approximates the inharmonicity of the note. In other words, our similarity routine is rather some likelihood calculation $A'(\beta, \hat{\beta}_s)$, discussed in detail in the next sections, between the measured inharmonicity β of the note and the predicted inharmonicity $\hat{\beta}_s$ of each of the candidate fretboard positions. This is an algorithmically simple approach, requiring merely inharmonicity estimation of the unknown note, inharmonicity trends of the strings of training guitars, and a likelihood measure. Additionally, this method is governed by the intuitive spatial constraint of inharmonicity

trajectories: notes played on a given string can exhibit only a set of inharmonicities, so classifying based on a spatially-informed metric, like proximity to a trajectory line, makes physical sense.

3.1.2 Inharmonicity Estimation

The first step in our system is estimation of the inharmonicity of an unseen note. The algorithm we employed was the median adjustive trajectories (MAT) method [?] proposed by Hodgkinson in 2009, and we present it here. At the core of this method is a recasting of the definition of inharmonicity in terms of any two partials and their indices. We can rearrange equation (2.1) to express the fundamental f_0 in terms of the frequency f_m of the m th partial and its index m as

$$f_0 = \frac{f_m}{m\sqrt{1 + \beta m^2}}. \quad (3.2)$$

Next, we can substitute (3.2) back into f_0 in equation (2.1) to obtain

$$f_k = \frac{k f_m}{m\sqrt{1 + \beta m^2}} \sqrt{1 + \beta k^2}. \quad (3.3)$$

Solving for β yields

$$\beta = \frac{(f_k \frac{m}{k})^2 - f_m^2}{k^2 f_m^2 - m^2 (f_k \frac{m}{k})^2}, \quad (3.4)$$

which defines the inharmonicity of a note in terms of an arbitrary two partials k and m and their corresponding frequencies f_k and f_m .

The MAT algorithm first pre-processes a note with the Complex Spectral Phase Evolution (CSPE) method [?] to refine the subsequent frequency estimates. The inharmonicity routine then begins by locating partials 1 and 2 in this refined spectrum based on a user-input estimate of the fundamental (slight mis-tunings are not a problem, since the algorithm refines the f_0 estimate with each iteration as will be seen). Searching for these partials is straightforward, since the skew effect of inharmonicity is low for small indices. From these locations, (3.4) is used for every

combination of the located partials (which for this first iteration, is only one) to obtain an array of β estimates. The median of the β array is returned as the composite β estimate of this iteration, and this median together with the locations of the partials are used in (3.2) to obtain an array of revised f_0 estimates that contains as many elements as partials under consideration. Finally, the median of this f_0 array and the median of the β array are used in equation (2.1) to predict the frequency of the next partial for which to search, which is the third in this case. At succeeding iterations, the lengths of these f_0 and β arrays grow longer (since both the number of partials, and therefore the number of combinations of partial pairs, increases), and the estimates which they contain grow increasingly more accurate. The algorithm is summarized in the flowchart in Figure 3.1.2, taken from [?]. MAT determines the ideal number of partials for which to search by finding the highest partial index that lies above the average level of the magnitude spectrum, and terminates once this partial index is reached. The result is an efficient and accurate inharmonicity estimation routine.

We apply the MAT estimation procedure to five Hann-windowed audio segments of length 2^{14} samples, beginning at the detected onset of a note and at four successive 100-ms intervals (i.e. 0 ms, 100 ms, 200 ms, 300 ms, and 400 ms after the note onset). This yields five inharmonicity estimates which we aggregate into one effective result using the frame aggregation process described in Section 3.1.5.

3.1.3 Log-inharmonicity

Next in our pipeline, we apply a log transformation to the estimated inharmonicities to make their variation with respect to MIDI pitch linear. This has the advantage of using simpler regression models for classification at later stages. If we reconsider equation (3.1) in terms of MIDI pitch number m instead of fret number n , we obtain

$$\beta(s, m) = \beta(s, m_{os}) 2^{\frac{m - m_{os}}{6}}, \quad (3.5)$$

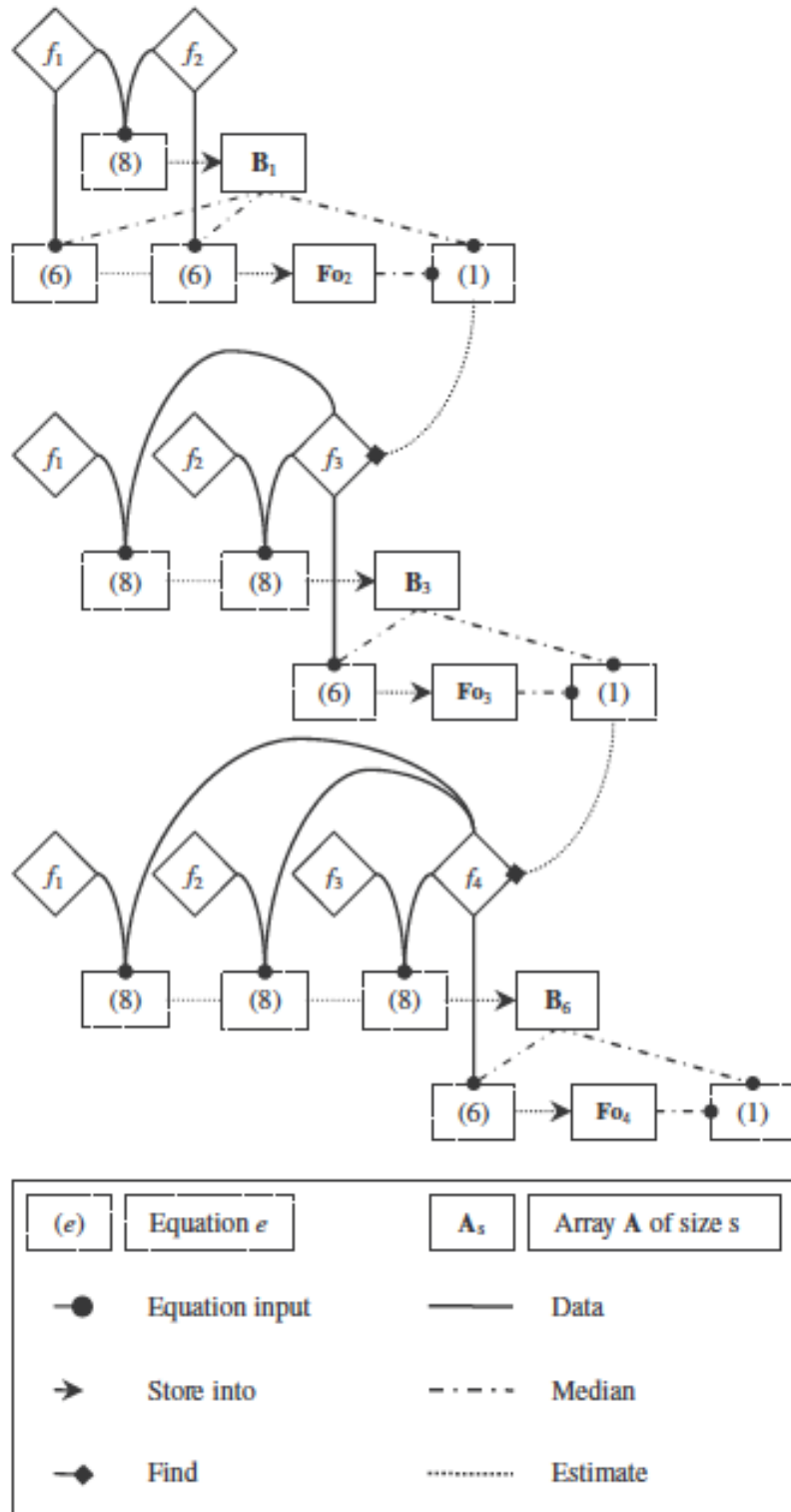


Figure 3.2: Median-adjustive-trajectory (MAT) method flowchart.

where m_{os} is the open-string pitch of string s . The exponential trajectory is maintained since incrementing both fret number and MIDI pitch number accomplishes an equivalent increase in fundamental frequency. Figure 3.3 illustrates the trajectories in terms of MIDI pitches.

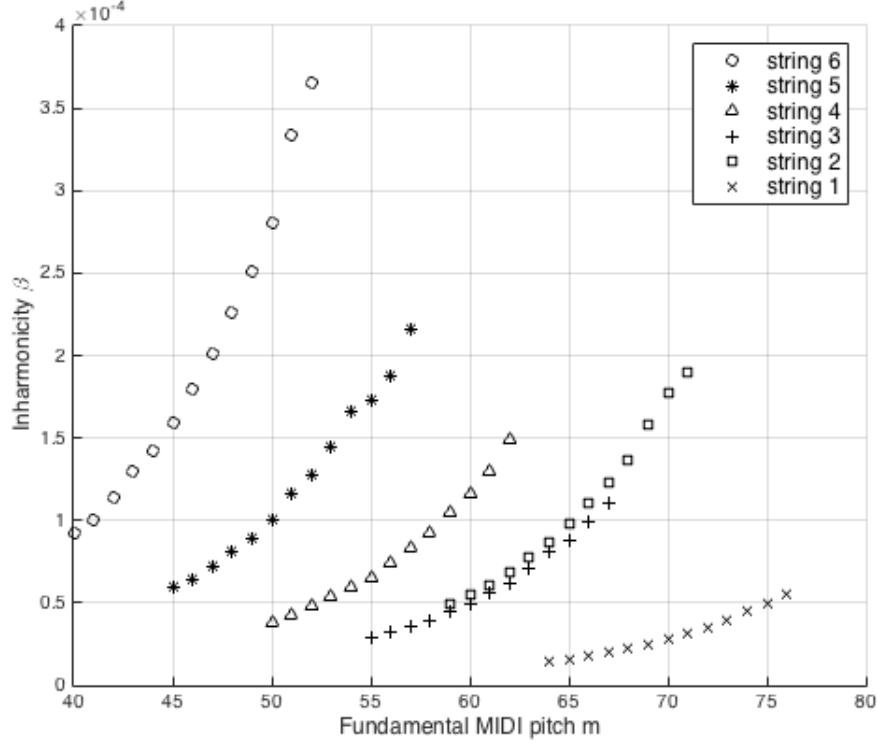


Figure 3.3: Inharmonicity estimates for Frets 0-12 on Strings 1-6 of RWC Acoustic Guitar AG111. Note the exponential trajectory as in Fig. 3.1 and the increased segregation since notes and their inharmonicities are now plotted against their fundamental instead of their fret number.

Now if we consider the log-inharmonicity $\beta_l(m)$, we see that

$$\beta_l(m) = \log_2 \beta(s, m) = \log_2 [\beta(m_{os}) 2^{\frac{m-m_{os}}{6}}] \quad (3.6)$$

$$\beta_l(m) = \log_2 \beta(m_{os}) + \frac{m - m_{os}}{6} \quad (3.7)$$

$$\beta_l(m) = (\log_2 \beta(m_{os}) - \frac{m_{os}}{6}) + (\frac{1}{6})m, \quad (3.8)$$

where string s has been dropped from the notation since we're looking only at variation along a

fixed string. Substituting w_0 for $\log_2 \beta(m_{os}) - \frac{m_{os}}{6}$ and letting $w_1 = \frac{1}{6}$, we obtain the familiar linear form

$$\beta_l(m) = w_0 + w_1 m, \quad (3.9)$$

showing us that the log-inharmonicity trajectory of a given string varies linearly with respect to MIDI pitch. See Figure 3.4. This relationship also holds for fret number, but fundamental pitch is a more likely and generic feature to have been estimated, and consequently is more relevant to consider here.

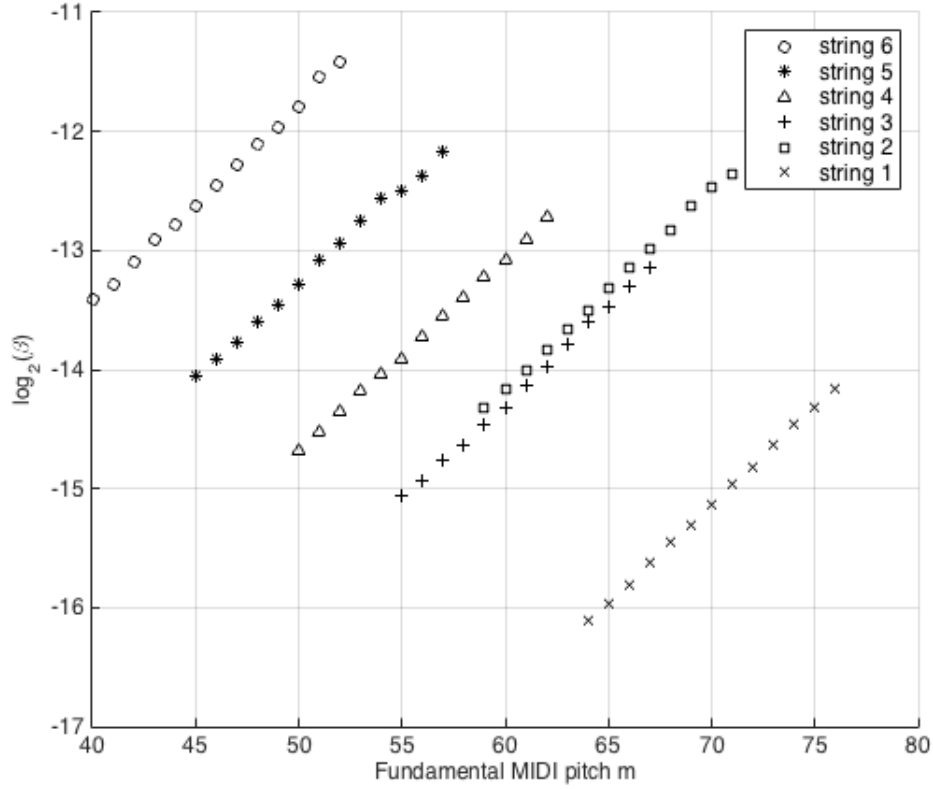


Figure 3.4: Log-inharmonicity estimates for same frets, strings, and guitar as Fig. 3.3.

3.1.4 Learning Inharmonicity Trajectories

We next need to train our system to understand the typical log-inharmonicity trajectories of guitar strings. There are two methods we use to do this. The first, which we call *theoretical* trajectory determination, is essentially the method employed by the existing PCT method [?]. For each guitar, we collect the six notes labeled with Fret 0 (i.e. played open-string) from its recordings, and estimate their inharmonicities. Our line trajectories are then simply those defined by the weights in equation (3.9) in terms of the open-string inharmonicities. The “learning” in this case is just the estimation of open-string inharmonicities; the lines themselves are simply extrapolations of the expected theoretical relationship between the log-inharmonicity of a string and its pitch.

The other method we use is linear regressions for obtaining *empirical* log-inharmonicity trajectories. Regressing log-inharmonicities of notes against their pitches could capture a more fitting trajectory than the strictly theoretical trajectories obtained from equation (3.9). In this method, we collect all notes with common string labels in our training data, and we perform linear regression of their log-inharmonicities β against their fundamental pitches m (in MIDI note number format). We do this for each string label. Let $s \in \{1, 2, 3, 4, 5, 6\}$ be the string label to which each training note is assigned, and N_s be the number of notes we have belonging to string label s . If we let $\mathbf{x}_s^{(i)} = [1, m_s^{(i)}]^T$ represent the i th note with fundamental pitch $m_s^{(i)}$ belonging to string s , and $\beta_s^{(i)}$ be the log-inharmonicity of the i th note belonging to string s , we can solve

$$\mathbf{w}_s = \arg \min_{\mathbf{w}} \sum_{i=1}^{N_s} (\beta_s^{(i)} - \mathbf{w}^T \mathbf{x}_s^{(i)})^2 \quad (3.10)$$

which yields the weight vector \mathbf{w}_s that minimizes the sum of squared error between the measured and predicted inharmonicities of the notes belonging to string s . Figure 3.5 illustrates the differences between the trajectory and regression approaches.

Occasional estimates from the inharmonicity measurement routine would be outliers that

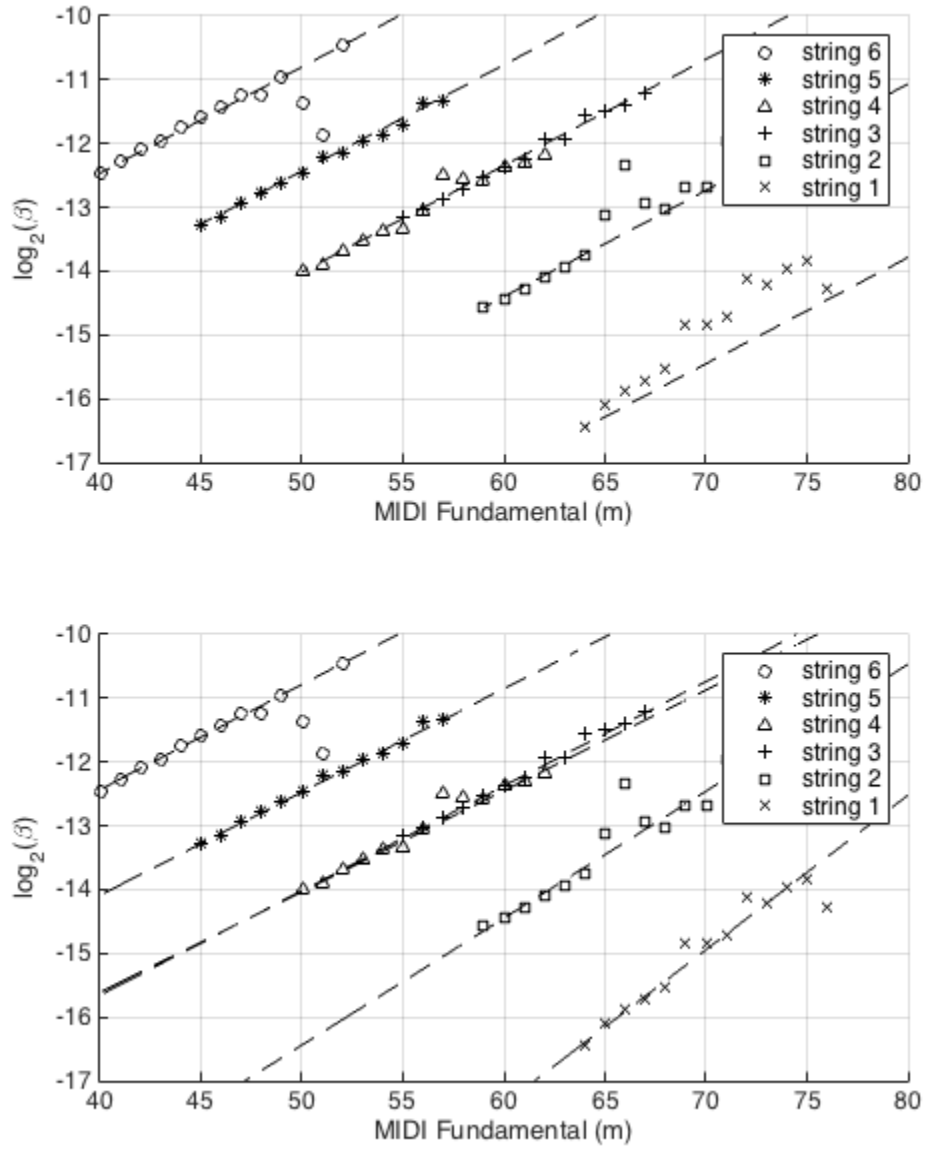


Figure 3.5: Comparison of string inharmonicity characterization approaches for our recorded Fender Telecaster. Top: *theoretical* trajectories. Note how each line begins precisely at the first marker of each string (i.e. the open-string note) and continues along a trajectory that ignores empirical inharmonicity estimates. String 1 most clearly shows this. Additionally, the trajectories for String 3 and String 4 are effectively co-linear and therefore indiscriminable. Bottom: regressions, or *empirical* trajectories. Each regression was obtained by minimizing the bisquare weighted error of the residuals of the string, so the lines more closely follow the empirical inharmonicity scatterplot estimates.

deviated substantially from the linear trajectory model, and which disproportionately skewed the regressions. For this reason, we rather performed a bisquare weighted linear regression, instead of the standard unweighted regression in (3.10), to empirically approximate the log-inharmonicities. Bisquare weighting is a robust linear regression procedure that minimizes a *weighted* sum of squares which de-emphasizes outliers. This provided us with higher-quality regressions that were more robust to stray inharmonicity measurements. An overview of the iterative procedure, implemented using the Matlab linear-model fitting routine ‘fitlm’, is presented below and is taken from the online documentation of the routine [?]. First, a regular least squares fit is obtained. Then, the standardized adjusted residuals are computed according to

$$u = \frac{r_i}{K s \sqrt{1 - h_i}}, \quad (3.11)$$

where r_i are the plain residuals, h_i are the leverages, $K = 4.685$ is a tuning constant, and s is the robust variance, which is equal to $M/0.6745$ where M is the median absolute deviation of the residuals. Next, the robust weights are computed according to

$$w_i = \begin{cases} (1 - u_i^2)^2, & |u_i| < 1 \\ 0, & |u_i| \geq 1 \end{cases} \quad (3.12)$$

The process is repeated, iteratively improving the weights until the fit converges.

3.1.5 String Classification

The next step is to use these learned string trajectories to classify unknown notes. Since these trajectories tell us how to expect the inharmonicities of different strings to vary with respect to pitch, determining the trajectory which best approximates the inharmonicity estimate of the unseen note appears to be a fitting procedure to ascertain the originating string. Specifically, we impose Gaussian distributions over the inharmonicity measurements of each string centered at

their trajectories, and evaluate them at the inharmonicities of unseen notes to determine the most likely string from which it came.

We can think of each inharmonicity trajectory as rather a trajectory of one-dimensional Gaussian log-inharmonicity distributions, whose means are the log-inharmonicity values of points on the line, and whose variances are those of the residuals of the regressions, σ_s^2 . In this manner, the probability distribution that characterizes the inharmonicity trajectory \mathbf{w}_s of string s at the fundamental pitch $m_0^{(i)}$ of the i th note $\mathbf{x}^{(i)} = [1, m_0^{(i)}]^T$ is defined as the normal distribution $\mathcal{N}(\mu, \sigma^2)$, with mean $\mu = \mathbf{w}_s^T \mathbf{x}^{(i)}$ and variance $\sigma^2 = \sigma_s^2$. Now to classify the n th unseen note $\mathbf{x}^{(n)}$, we simply measure its log-inharmonicity $\beta^{(n)}$ and obtain the probability that this log-inharmonicity was observed given each string s according to

$$P(\mathbf{x}^{(n)}, \beta^{(n)} | s) = \mathcal{N}(\beta^{(n)} | \mathbf{w}_s^T \mathbf{x}^{(n)}, \sigma_s^2), \quad (3.13)$$

and return the predicted string $\hat{s}^{(n)}$ that maximizes this probability,

$$\hat{s}^{(n)} = \arg \max_s P(\mathbf{x}^{(n)}, \beta^{(n)} | s). \quad (3.14)$$

Another conceivable decision function is classification based on the residual distance of an unseen note to the string trajectories. Instead of selecting the string whose trajectory maximizes the probability of inharmonicity observation, one could select the string whose trajectory exhibits minimum residual with the inharmonicity measurement of the note. Indeed, this worked well, though experiments comparing this “hard” residual classifier versus the “soft” probabilistic classifier found that the probabilistic method performed slightly better, so we used this for our experiments in Chapter 4.

Again, outlier inharmonicity estimates would throw off this classification scheme. To combat this, we employed a variant of frame aggregation, a refinement technique used by [?]. In frame aggregation, inharmonicities of multiple audio windows are estimated and classified individually,

then the multiple results are aggregated into a compound final decision. In essence, it is a form of averaging that relies on the theoretically low probability that consecutive inharmonicity estimates will all be outliers; the more probable stable estimates will contribute to a more reasonable aggregate string decision. For each frame r in an R -frame aggregation of a note, we obtain the string-wise probabilities $\mathbf{p}_r = [p_1, p_2, p_3, p_4, p_5, p_6]^T$ of its log-inharmonicity estimate for all possible strings. Then, we simply sum the probabilities \mathbf{p}_r over all frames and select the string whose aggregate likelihood is maximal.

The result is a slight boost in classification accuracy. For our system, we used five overlapping Hann-windowed frames, each 2^{14} samples long, which start at 100-ms intervals after the detected onset of a note. Onsets were obtained through automated detection for convenience, the details of which are beyond the scope of this work but are cited for reference [? ?]. This frame-aggregation procedure is possible only because the sustains of the RWC notes are all quite long, and this is, of course, a limitation. The entire analysis spans approximately three-fourths of a second, which implies that the analyzed music passage cannot exceed rhythms greater than quarter notes at 80 bpm – this is an admittedly unrealistic requirement of guitar audio, notorious for featuring licks played with rapidity that exceeds eighth and sixteenth notes at considerably higher tempos. We therefore emphasize the theoretical, proof-of-concept nature of this work.

3.1.6 Tablature Conversion and Refinement

The final step is transformation of the string classifier output into tablature notation. Because the classifier outputs are string numbers, we still need to infer the fret positions of notes. This is a trivial task provided we know the tuning $\mathbf{t} = [m_1, m_2, m_3, m_4, m_5, m_6]^T$ of the unknown guitar, where m_s is the MIDI pitch number of the open note on string s . Because we already have an estimate of the MIDI pitch $m^{(n)}$ of the n th unknown note, simply taking the difference between $m^{(n)}$ and the open pitch m_s of its assigned string yields the fret on which the note was played; each fret on a guitar increases the pitch of the string by one half-step, or equivalently the MIDI

pitch number by one integer.

We further refine the fretboard estimates with “plausibility filtering” [?], in which we reject and reassign those string decisions that are implausible given the constraints and assumptions of the data. If our systems assigns to a note a string on which the corresponding fretted position is negative, a mistake has clearly been made since frets cannot possibly be lower than 0 (the open-string). Similarly, if our system assigns to a note a string on which the corresponding fret is greater than 12, we know this is an error as well because our guitar recordings feature only Frets 0-12. When either of these situations arise, we reject the string classification decision and select the next most probable string. If the succeeding string is also implausible, the process is iterated until all six strings have been considered.

3.2 Baseline Bayesian Classifier

3.2.1 Motivation

The hypothesis of our first method is that exploitation of the linear trajectories of log-inharmonicities is helpful for transcription. In order to objectively evaluate this, we need for comparison the results of a baseline classifier which *do not* leverage the inharmonicity trajectories. This method, which we call simply Bayesian classification of fretboard position, serves as this benchmark requirement.

3.2.2 Inharmonicity Estimation

As before, the first step in this method is measurement of the inharmonicities of notes. We again use median-adjustive-trajectories (MAT) to estimate this crucial feature. MAT is explained in detail in Section 3.1.2.

3.2.3 Fretboard-Position Inharmonicity Distributions

Next, we need to capture the inharmonicity distribution of each fretboard position f . For each of the $F = 78$ positions ($13 \text{ notes} \times 6 \text{ strings}$), we calculate the median $\tilde{\beta}_f$ and variance σ_f^2 of its inharmonicities, and use these to parameterize a Gaussian probability density $\mathcal{N}_f(\tilde{\beta}_f, \sigma_f^2)$. The median was chosen to lessen the influence of outliers, and was found to produce slightly better results than the mean.

3.2.4 Fretboard Position Classification

Now that each fretboard position f has been characterized as a probability distribution \mathcal{N}_f of inharmonicities, we can classify an unseen note. With the interpretation of inharmonicity as a random variable, we measure the inharmonicity $\beta^{(i)}$ of the i th unseen note $\mathbf{x} = [1, m_0^{(i)}]^T$ with fundamental MIDI pitch $m_0^{(i)}$ and simply evaluate the probability $P_f(\beta^{(i)}|f) = \mathcal{N}_f(\beta^{(i)}|\tilde{\beta}_f, \sigma_f^2)$ of it having been generated by the Gaussian distribution \mathcal{N}_f . We narrow our search space $f \in \{1, 2, 3, \dots, F-1, F\}$ of possible fretboard positions by considering just the positions $f \in \{f_{m,1}, f_{m,2}, \dots\}$ that agree with the fundamental frequency of the unseen note. For standard-tuning, this is a maximum of only three positions. To classify a note, we select the candidate fretboard position which maximizes the likelihood of the measured inharmonicity of the note, just as in equation (3.14):

$$\hat{f} = \arg \max_{f \in \{f_{m,1}, f_{m,2}, \dots\}} P(\beta^{(i)}|f). \quad (3.15)$$

We also perform frame aggregation with this classification method, using the same specifications as discussed in Section 3.1.5: five overlapping Hann-windowed audio segments of length 2^{14} samples starting at the onset of a test note and at successive 100-ms intervals thereafter. These windows are individually analyzed to produce five inharmonicity measurements. Each measurement is used in equation (3.15), and the individual probabilities for candidate fretboard positions are summed across the five measurements. The position with maximum aggregate probability

is selected as the classification output. Again, the unrealistic tempo limitation of this procedure must be recognized – quarter notes at 80 bpm is a restrictive upper bound for the rhythm of a guitarist.

3.3 Tuning Compensation Feature

After transforming our data into log-inharmonicity space in the first method, or after estimating the probability distributions of the inharmonicities of each fretboard positions in the second method, we can apply a tuning compensation feature step if the test guitar is known to be in alternate-tuning. In this section, we derive the straightforward adjustment and explain how it is incorporated into our system, focusing on the context of the inharmonicity trajectory method.

Though standard tuning is the most common pitch configuration of six-string guitars, there exist numerous other tunings in which performers often play. Aside from altering the musicality of the instrument, alternate tunings complicate the transcription process by introducing uncertainty about the open-string pitches. They distort inharmonicity-based methods, since the tensions of alternately-tuned guitar strings differ from those in standard-tuned strings, thereby affecting estimation of the inharmonicity.

Current string classification systems do not address this degree of freedom. PCT [?] is technically tuning-invariant, but requires access to recordings of the open strings of alternately-tuned test guitars. Other supervised learning systems [? ? ?] are trained and evaluated only on standard-tuning guitars.

A possible approach to augment inharmonicity-based string classifiers with tuning-invariant performance is to simply introduce a scaling factor on the expected inharmonicity, or equivalently an additive factor on the expected log-inharmonicity. The fundamental frequency f_0 of an ideal vibrating string is related to its tension T according to

$$f_0 = \frac{1}{2L} \sqrt{\frac{T}{\mu}} \quad (3.16)$$

where L is string length and μ is its density. From this, we can see that

$$T \propto f_0^2. \quad (3.17)$$

Recognizing that the equivalent change in frequency for a change in fundamental pitch of Δm semitones is $2^{\frac{\Delta m}{12}}$, we see that the proportional change in tension (with all other factors constant) is

$$T \propto 2^{\frac{\Delta m}{6}} f_0^2. \quad (3.18)$$

The resulting log-inharmonicity of this new open-string note with pitch $m'_{os} = m_{os} + \Delta m$, with original open-string pitch m_{os} , is therefore

$$\log_2 \beta(m'_{os}) = \log_2 \left[\frac{\pi^3 Q d^4}{64 T l^2} (2^{-\frac{\Delta m}{6}}) \right]. \quad (3.19)$$

$$\log_2 \beta(m'_{os}) = \log_2 \frac{\pi^3 Q d^4}{64 T l^2} - \frac{\Delta m}{6}. \quad (3.20)$$

Substituting (3.20) into the log-inharmonicity term of w_0 from equation (3.9), we can get the new intercept term w'_0 :

$$w'_0 = \log_2 \beta(m'_{os}) - \frac{m'_{os}}{6} \quad (3.21)$$

$$w'_0 = \left(\log_2 \frac{\pi^3 Q d^4}{64 T l^2} - \frac{\Delta m}{6} \right) - \frac{m_{os} + \Delta m}{6} \quad (3.22)$$

$$w'_0 = \log_2 \frac{\pi^3 Q d^4}{64 T l^2} - \frac{m_{os}}{6} - \frac{\Delta m}{3} \quad (3.23)$$

$$w'_0 = w_0 - \frac{\Delta m}{3}. \quad (3.24)$$

The slope coefficient $w_1 = -\frac{m}{6}$ is independent from Δm , so the affected log-inharmonicity trajectory of this alternately-tuned string is therefore

$$\beta'(m) = w_0 + w_1 m - \frac{\Delta m}{3}. \quad (3.25)$$

In words, equation (3.25) tells us that the effect of an alternate tuning on a string is simply introduction of a bias term $\frac{-\Delta m}{3}$ to the log-inharmonicity, where Δm is the semitone deviation from standard tuning. We should thus be able to compensate our predictive trajectories to arbitrarily-tuned guitars, provided we are given the alternate tuning in which the unseen notes are being performed. As can be seen in Figure 3.3, when one string on the standard-tuned electric guitar is modified using equation (3.25), the predictive alternate-tuning regression aptly captures the trend of the alternate-tuning inharmonicity. This is a straightforward modification for any inharmonicity-exploiting classifier, and we report our results with this modification in the next chapter.

3.4 System Overviews

Below in Figure 3.7, we summarize the operation of our inharmonicity trajectory system. Inharmonicity estimates are obtained for the notes of our training guitars, then we represent these estimates in log-inharmonicity space with respect to MIDI pitch to obtain linear trends. We then check if the tunings of the evaluation guitars are non-standard, and if so, we appropriately compensate the weight vectors that define the linear trajectories. Next, we perform classification on test recordings by assigning to unknown notes the strings whose Gaussian log-inharmonicity trajectories maximize the likelihood of the log-inharmonicities of the notes. Finally, tablature conversion and refinement is performed with plausibility and constraint filtering.

Our baseline Bayesian fretboard-position classifier is summarized in Figure 3.8. It crucially differs from the inharmonicity trajectory method in that it does not exploit any aspect of the trajectory at all; it simply performs likelihood-maximization using the individual fretboard inharmonicity distributions, providing a relevant benchmark of transcription performance. First, inharmonicity estimates for the training guitar notes are obtained, and normal distributions are fit to the inharmonicity variation of each fretboard positions. Then, if the tuning of the test guitar is non-standard, the distributions can be compensated appropriately. Notes from the test guitar

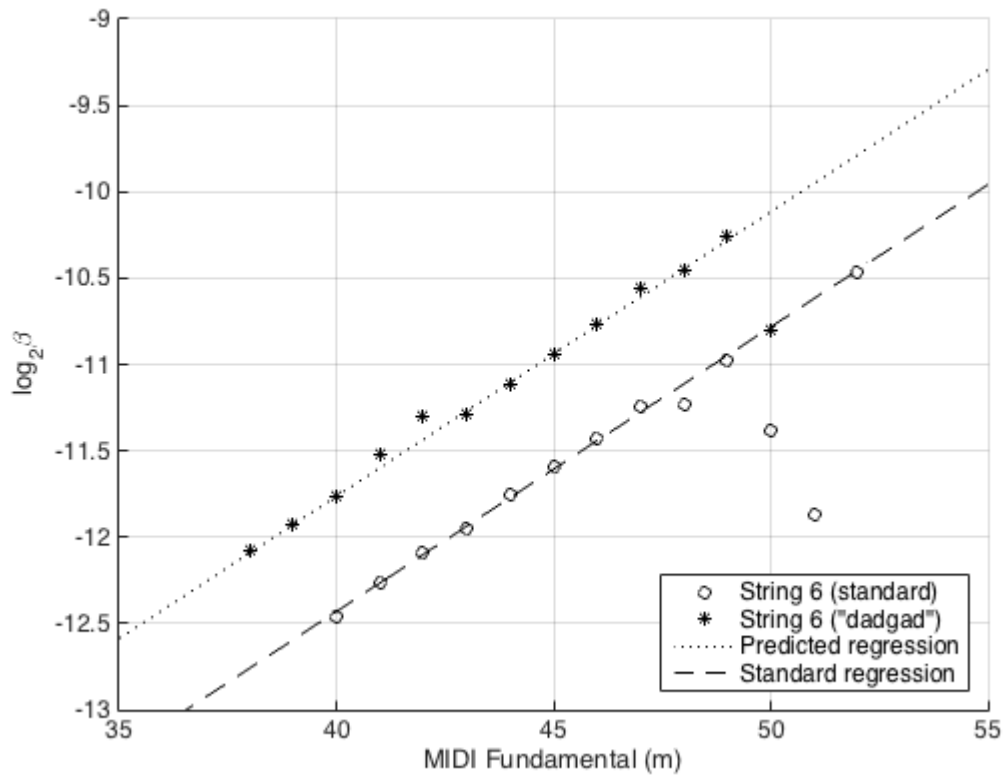


Figure 3.6: Effect of alternate tuning on log-inharmonicity trajectories of an electric guitar. The circles are log-inharmonicity estimates of Frets 0-12 on String 6 of RWC Electric Guitar EG131 in standard tuning, and the dashed line is their bisquare-weighted regression. The asterisks are log-inharmonicity estimates for the same frets, string, and guitar recorded two semitones down. The dotted line passing through them is the *predicted* regression obtained by applying equation (3.25) to the dashed line.

are classified by measuring their inharmonicities and, as before, selecting the fretboard positions that maximize the likelihood of their observation. Tablature refinement is built-in by considering as candidate fretboard positions only notes whose fundamental pitches match those of the unseen notes.

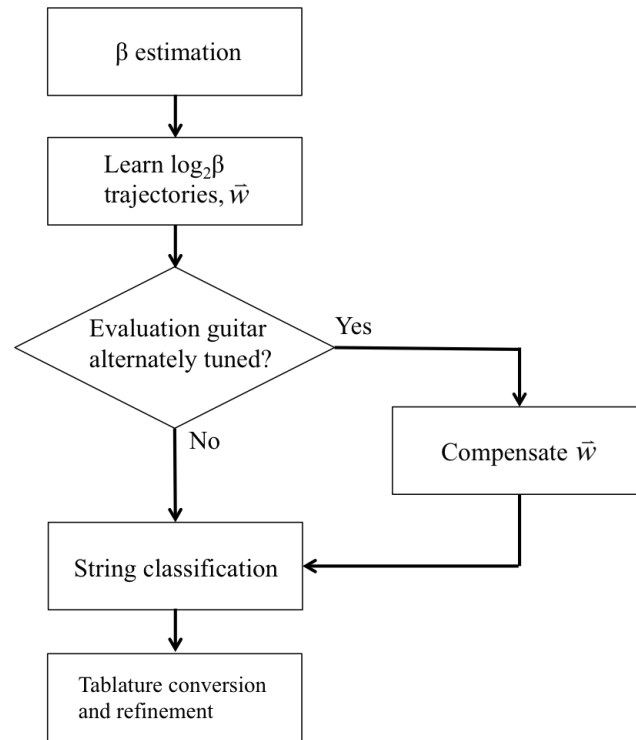


Figure 3.7: System overview of tablature transcription with log-inharmonicity trajectories using probabilistic classification.

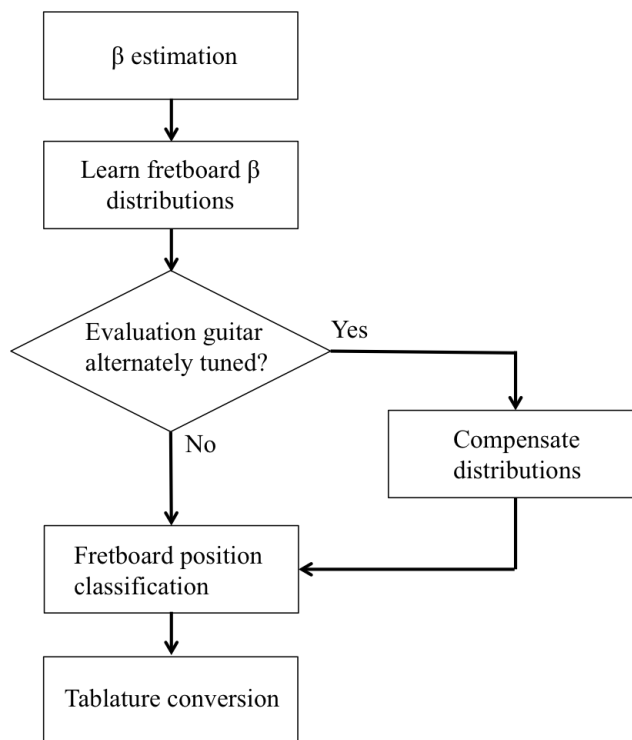


Figure 3.8: System overview of tablature transcription with Bayesian fretboard-position classification.

Chapter 4: Experiments and Results

4.1 RWC Evaluation

We used subsets of the Real World Corpus Music Instrument Database (RWC-MDB-I) [?] and personal recordings of an electric and acoustic guitar to conduct our experiments.

Our RWC subset comprised nine guitar recordings (three classical, three acoustic, and three electric), each of which were performed twelve times exhibiting various permutations of particular musicality parameters (playing style, dynamic level, pickup selection). For our work, we used only six of the twelve recordings for each of the electric guitars, as the omitted ones featured musicality parameters like vibrato and palm-muting that were not featured in the classical and acoustic guitar sets and which made inharmonicity estimation difficult. At the time of the writing of this thesis, we were not certain whether [?] also used this smaller subset of electric guitars, so we avoid comparison of our electric guitar results to theirs; we accordingly draw comparisons for electric guitar results only between our baseline and our regression method. The RWC recording performances themselves were simply clean, isolated, monophonic enumeration of every fret (from open-string to 12th fret) on every string (from string 6 to string 1), constituting 78 total note plucks per audio file. The resolution is 16 bits per sample, at 44.1 kHz. Labels of strings and frets are thus obtained by their location of occurrence in the recording. The classical guitars recorded were a Stafford, a Sakurai Kohno Professional-J, and a Yuichi Imai YJ-II; the three acoustic guitars captured were an Ovation, a Yamaha APX, and Yairi WY1; the electric guitars featured were a Fender Stratocaster, an Aria PE, and an Ibanez Artcore.

We conduct our experiments on all nine RWC guitars at three levels of context, which comprise the headings of Table 4.1: guitar-specific, guitar-averaged, and guitar-independent. For guitar-specific trials, we train and test the system on the same guitar using three-fold cross-validation with test folds that are one-third of the guitar recordings. The results of the folds are averaged together for composite transcription errors. For guitar-averaged trials, we train the system on all recordings of the two guitars in the same class as the test guitar, as well as a two-thirds training portion of the test guitar, and then evaluate on the remaining third of the test guitar recordings. Again, three-fold cross-validation is used to iterate through the portions of the test recordings. For guitar-independent trials, the system never sees any of the recordings of the test guitar before evaluation. We train the system on two of the guitars in the same class as the test guitar, and test on all recordings of the remaining one. In all cases, Table 4.1 reports error probabilities, as this was the transcription metric used in [?].

In the “Bayes” columns of each heading in Table 4.1, we report transcription results using the baseline Bayesian classifier we developed. To review, Gaussian distributions were fit to the inharmonicity measurements of every fretboard position in the training data. Then, notes were classified by selecting the fretboard positions that most probably produced their measured inharmonicities, according to the learned Gaussian likelihoods.

We next report string identification results (taken from [?]) using the PCT method under the “PCT” columns of Table 4.1. As discussed in Chapter 2, this method uses estimates of the open-string inharmonicities of the training guitar audio to predict *expected* inharmonicities of the candidate fretboard positions of an unknown note. The degree of similarity between the measured inharmonicity of an unknown note and that of its candidate fretboard positions is assessed by tallying their coincident partials. The candidate fretboard position that produces the highest tally is returned as the transcribed tablature. Note that no results for the “Guitar-independent” column exist for this method, as no results were reported in [?] for this scenario.

Finally, we evaluate our trajectory classification method, reporting error probabilities using empirical trajectories (i.e. regressions) under the “Regr.” columns of Table 4.1. We performed

bisquare weighted regression of log-inharmonicity estimates of training guitar audio against their pitches. String classification was performed on the test notes by selecting the regressions whose imposed Gaussian p.d.f.s maximized the likelihood of the inharmonicities of the notes, as discussed in Chapter 3. We also evaluated our *theoretical* trajectory method, in which regressions are not computed but likelihood maximization is still used for classification. Error probabilities were similar to those of our *empirical* trajectory (i.e. regression) method, however, so we omitted theoretical trajectory results for conciseness. Table 4.1 is dense, so we distill its central results in the next paragraphs during our discussions of its statistical significance.

Transcription Error Probabilities									
	Guitar-specific			Guitar-averaged			Guitar-independent		
Guitar	Bayes	PCT	Regr.	Bayes	PCT	Regr.	Bayes	PCT	Regr.
Classical CG091	0.02	0.04	0.03	0.13	0.09	0.03	0.14	–	0.03
Classical CG092	0.15	0.00	0.06	0.16	0.01	0.06	0.13	–	0.03
Classical CG093	0.08	0.01	0.04	0.14	0.00	0.06	0.14	–	0.06
Overall CG:	0.08	0.02	0.04	0.14	0.03	0.05	0.14	–	0.05
Acoustic AG111	0.00	0.00	0.03	0.01	0.06	0.03	0.02	–	0.03
Acoustic AG112	0.02	0.00	0.03	0.03	0.22	0.06	0.13	–	0.13
Acoustic AG113	0.03	0.00	0.04	0.06	0.04	0.06	0.10	–	0.07
Overall AG:	0.02	0.00	0.03	0.03	0.11	0.05	0.08	–	0.08
Overall CG, AG:	0.05	0.01	0.04	0.09	0.07	0.05	0.11	–	0.06
Electric EG131	0.22	–	0.18	0.21	–	0.18	0.24	–	0.23
Electric EG132	0.15	–	0.10	0.15	–	0.17	0.17	–	0.21
Electric EG133	0.06	–	0.09	0.20	–	0.19	0.22	–	0.27
Overall EG:	0.14	–	0.12	0.19	–	0.18	0.21	–	0.24
Overall CG, AG, EG:	0.08	–	0.07	0.12	–	0.09	0.14	–	0.12

Table 4.1: Error probability comparisons between the baseline classifier (Bayes), the existing partial coincidence tally method (PCT), and our proposed probabilistic regression classifier (Regr.). PCT results are taken from [?]; we’ve excluded error due to f_0 estimation in their results. Note that PCT results for electric guitars are intentionally excluded to avoid potentially unfair comparison, while PCT results for the “Guitar-independent” scenario are not available.

We performed paired-sample t -tests to assess whether differences in error rates between the existing PCT method and each of our novel methods were statistically significant at the $\alpha = 0.05$ level. We computed the difference between the error probabilities of the PCT method and the

method-under-comparison (either Bayes or Regression) for each row in Table 4.1 to obtain μ_d , the mean of the distribution of the differences. Because we did not have access to the trial-by-trial transcription results of PCT, we approximated the standard deviation σ_d of the distribution of the differences by the square root of the average of the variances of the methods $\sigma_d = \sqrt{(\sigma_1^2 + \sigma_2^2)/2}$ according to the Bernoulli approximation, in which $\sigma_x^2 = p_x(1-p_x)$ and the success parameter p_x is simply the difference of 1 and the reported error probability of method x , i.e. $p_x = 1 - \mu_x$. We calculated the t -score of each comparison according to $t = \mu_d/(\sigma_d/\sqrt{n})$, with $n = 78 \text{ notes} \times 12 \text{ recordings} = 936$ classified instances over which the error probabilities were averaged ($n = 468$ for the fewer recordings of the electric guitars, $n = 2808$ for the overall error measures by guitar type, and $n = 7020$ for the total overall error measures), then summed the right and left tails of the c.d.f of the student t -distribution at $\pm t$. This gives us the probability, or p -value, of observing the performance differences of the classifiers due to chance given that the null hypothesis of zero-mean performance differences is true. Table 4.2 shows a subset of Table 4.1 in which we highlight transcription results from the t -tests which yielded p -values less than $\alpha = 0.05$, allowing us to reject the null hypothesis in favor of the alternative that the performance differences of the classifiers do not have zero-mean, i.e. that the highlighted results differ significantly from PCT. The bottom-most “Overall CG, AG” row shows that our regression method improves upon PCT’s performance under the “Guitar-averaged” scenario to a statistically significant degree across all classical and acoustic guitars.

We next performed two-tailed McNemar tests on results of our regression and baseline methods to assess the validity of the null hypotheses that both classifiers performed equally well for each RWC evaluation guitar at the $\alpha = 0.05$ significance level. (See Table 4.3). Transcripts of the $78 \text{ notes} \times 12 \text{ recordings} = 936$ string assignments of our two classifiers for each evaluation guitar (468 assignments for the fewer recordings of the electric guitars, 2808 for the overall performance by guitar type, and 7020 for the total overall performance) along with their true labels were examined with the Matlab ‘testcholdout’ function to provide p -values. Most notably in the bottom “Overall CG, AG, EG” row, our regressions statistically significantly out-

Transcription Error Probabilities - Significance Comparison w/ PCT									
	Guitar-specific			Guitar-averaged			Guitar-independent		
Guitar	Bayes	PCT	Regr.	Bayes	PCT	Regr.	Bayes	PCT	Regr.
Classical CG091	0.02	0.04	0.03	0.13	0.09	0.03	–	–	–
Classical CG092	0.15	0.00	0.06	0.16	0.01	0.06	–	–	–
Classical CG093	0.08	0.01	0.04	0.14	0.00	0.06	–	–	–
Overall CG:	0.08	0.02	0.04	0.14	0.03	0.05	–	–	–
Acoustic AG111	0.00	0.00	0.03	0.01	0.06	0.03	–	–	–
Acoustic AG112	0.02	0.00	0.03	0.03	0.22	0.06	–	–	–
Acoustic AG113	0.03	0.00	0.04	0.06	0.04	0.06	–	–	–
Overall AG:	0.02	0.00	0.03	0.03	0.11	0.05	–	–	–
Overall CG, AG:	0.05	0.01	0.04	0.09	0.07	0.05	–	–	–

Table 4.2: Error probability comparisons. Results from our baseline (“Bayes”) and regression (“Regr.”) methods which differ from PCT to a statistically-significant ($\alpha = 0.05$) degree are shaded in gray. Electric guitar trials and “Guitar-independent” scenarios are omitted since comparisons in those cases are not available.

perform the baseline when averaging across all RWC guitars types. We also found that for the “Guitar-specific” scenario, our regressions improved on CG2, CG3, and EG2, while our baseline performed better on AG1. All other differences under this heading were insignificant with p -values greater than 0.05. For the “Guitar-averaged” scenario, regressions excelled at all CGs, the baseline excelled at AG1 and AG2, and all other differences were insignificant. For the “Guitar-independent” scenario, the improvements of the regression method for CG1, CG2, CG3, and AG3 as well as the superiority of the baseline method for AG1, EG2, and EG3 were all statistically significant.

Also displayed in Tables 4.4, 4.5, and 4.6 are the *string-wise* F1-scores using our regression-based classifier for the “Guitar-averaged” scenario. The F1-score is the harmonic mean of precision and recall, and is a commonly used classification metric elsewhere, so we use it here and wherever comparison with [?] is absent. The left-hand columns of these tables denote the string number of each row, and the right-hand columns show the row-wise average of the three guitars in the class. We present this information to highlight the contributions of the individual strings to the overall transcription accuracy.

Transcription Error Probabilities - Novel Methods Significance Comparison									
	Guitar-specific			Guitar-averaged			Guitar-independent		
Guitar	Bayes	PCT	Regr.	Bayes	PCT	Regr.	Bayes	PCT	Regr.
Classical CG091	0.02	–	0.03	0.13	–	0.03	0.14	–	0.03
Classical CG092	0.15	–	0.06	0.16	–	0.06	0.13	–	0.03
Classical CG093	0.08	–	0.04	0.14	–	0.06	0.14	–	0.06
Overall CG:	0.08	–	0.04	0.14	–	0.05	0.14	–	0.05
Acoustic AG111	0.00	–	0.03	0.01	–	0.03	0.02	–	0.03
Acoustic AG112	0.02	–	0.03	0.03	–	0.06	0.13	–	0.13
Acoustic AG113	0.03	–	0.04	0.06	–	0.06	0.10	–	0.07
Overall AG:	0.02	–	0.03	0.03	–	0.05	0.08	–	0.08
Electric EG131	0.22	–	0.18	0.21	–	0.18	0.24	–	0.23
Electric EG132	0.15	–	0.10	0.15	–	0.17	0.17	–	0.21
Electric EG133	0.06	–	0.09	0.20	–	0.19	0.22	–	0.27
Overall EG:	0.14	–	0.12	0.19	–	0.18	0.21	–	0.24
Overall CG, AG, EG:	0.08	–	0.07	0.12	–	0.09	0.14	–	0.12

Table 4.3: Error probability comparisons. Superior results in trials with statistically-significant ($\alpha = 0.05$) performance differences are shaded in gray. PCT results omitted to highlight performance differences between our baseline and regression methods.

Transcription F1-scores				
String No.	CG1	CG2	CG3	Overall CG
1	0.96	0.93	0.96	0.95
2	0.92	0.90	0.91	0.91
3	1.00	0.95	0.96	0.97
4	0.96	0.93	0.91	0.93
5	1.00	0.96	0.93	0.96
6	1.00	0.98	1.00	0.99

Table 4.4: String-wise F1-scores for the RWC classical guitars, using regression classification for the “Guitar-averaged” scenario.

Transcription F1-scores				
String No.	AG1	AG2	AG3	Overall AG
1	0.96	0.96	0.96	0.96
2	0.94	0.80	0.87	0.87
3	0.91	0.89	0.85	0.88
4	1.00	1.00	0.99	1.00
5	1.00	1.00	0.99	1.00
6	1.00	1.00	1.00	1.00

Table 4.5: String-wise F1-scores for the RWC acoustic guitars, using regression classification for the “Guitar-averaged” scenario.

Transcription F1-scores				
String No.	EG1	EG2	EG3	Overall EG
1	0.96	0.94	0.96	0.95
2	0.94	0.95	0.72	0.87
3	0.47	0.42	0.42	0.44
4	0.66	0.72	0.79	0.72
5	0.85	0.92	1.00	0.92
6	0.96	0.98	1.00	0.98

Table 4.6: String-wise F1-scores for the RWC electric guitars, using regression classification for the “Guitar-averaged” scenario.

4.2 Tuning Compensation

For the next experiment, we also recorded a Fender Telecaster (electric guitar) and a Takamine G-series (acoustic guitar) at 16 bits and 44.1 kHz in the same vein as the RWC database: string and fret enumeration, totaling 78 notes per recording. The string gauge of the acoustic guitar was 0.12 inches, while that of the electric guitar was unknown. We captured various tunings: standard, “DADGAD” (in which the open pitches of String 6 through String 1 are given by the respective letters in name), “WSU” (whole-step up), and “WSD” (whole-step down). The Fender, which was recorded directly into our audio interface, was performed with a thick plectrum at center pickup orientation (both bridge and neck pickups engaged) and at similar moderate dynamic levels throughout. We recorded the Takamine in an isolated acoustic chamber with a large diaphragm condenser microphone (an MXL 990) placed 12 inches away from the twelfth fret, and we also played it with a thick plectrum and at constant dynamic levels. Analog-to-digital hardware included a Focusrite Saffire Pro 24 and a 2011 Macbook Pro running Logic. No other musicality parameters were captured; we focused on tuning variance here.

As a reference for those unfamiliar with the selected tunings, String 6 through String 1 are tuned to E2, A2, D3, G3, B3, and E4 in standard tuning. In so-called “DADGAD” tuning, the open-string pitches are set to D2, A2, D3, G3, A3, and D4, as these open pitches concatenated together form its name. Whole-step up and whole-step down are tunings in which each of the standard-tuned strings are respectively incremented or decremented by two semitones. For whole-step down, this results in D2, G2, C3, F3, A3, and D4. For whole-step up, we have F#2, B2, E3, A3, C#4, and F#4.

We learned inharmonicity regressions from the recordings of the standard-tuned RWC electric and acoustic guitar recordings, and predicted string classes on our alternate-tuned Fender Telecaster and Takamine G-series recordings, respectively. We evaluated string-wise F1-score performance both with and without the tuning compensation factor in our system discussed in

Chapter 3. Tuning compensation results are denoted by the “comp.” column, and original results without compensation are denoted by the “orig.” column. See Table 4.7. For reference, we also report transcription performance on our personal guitars in standard tuning, in the second column from the left. We highlight in gray the results which are statistically significant at the $\alpha = 0.05$ level, determined by a McNemar test. Highlighted cells under ‘orig.’ columns indicate that the performance degradation due to the corresponding tuning is significantly different from the standard-tuned performance (‘Standard’ left-most heading). Highlighted cells under ‘comp.’ columns indicate that the performance gain achieved by our compensation system is significantly different than its uncompensated counterpart.

Alternate-tuning Transcription F1-scores (Electric)							
	Standard	“DADGAD”		“WSU”		“WSD”	
String No.	orig.	orig.	comp.	orig.	comp.	orig.	comp.
1	0.96	0.96	0.96	0.87	0.96	0.96	0.96
2	0.96	0.92	0.92	0.70	0.96	0.96	0.96
3	0.50	0.48	0.43	0.65	0.50	0.50	0.50
4	0.76	0.76	0.74	0.38	0.76	0.76	0.76
5	1.00	1.00	0.87	0.56	1.00	1.00	1.00
6	1.00	1.00	1.00	1.00	1.00	1.00	1.00
Overall:	0.86	0.85	0.82	0.69	0.86	0.86	0.86

Table 4.7: String-wise F1-scores for our Fender Telecaster at various tunings. Regressions trained on recordings of all three RWC electric guitars. WSU: whole-step up (from standard); WSD: whole-step down (from standard); orig.: no tuning compensation used in this trial; comp.: tuning compensation used in this trial. The “Overall” row displays average F1-score of each column. Statistically significant ($\alpha = 0.05$) results highlighted in gray.

Alternate-tuning Transcription F1-scores (Acoustic)							
	Standard	“DADGAD”		“WSU”		“WSD”	
String No.	orig.	orig.	comp.	orig.	comp.	orig.	comp.
1	0.96	0.87	0.92	0.96	0.96	0.92	0.96
2	0.82	0.70	0.63	0.44	0.82	0.67	0.92
3	0.96	0.93	1.00	0.74	0.96	0.00	0.96
4	0.87	0.79	0.79	0.90	0.87	0.44	0.93
5	1.00	1.00	1.00	0.82	1.00	0.81	1.00
6	1.00	1.00	1.00	0.96	1.00	1.00	1.00
Overall:	0.93	0.88	0.89	0.80	0.93	0.64	0.96

Table 4.8: String-wise F1-scores for our Takamine G-series at various tunings. Regressions trained on all recordings of all three RWC acoustic guitars. WSU: whole-step up (from standard); WSD: whole-step down (from standard); orig.: no tuning compensation used in this trial; comp.: tuning compensation used in this trial. The “Overall” row displays average F1-score of each column. Statistically significant ($\alpha = 0.05$) results highlighted in gray.

Chapter 5: Discussion

In this chapter we begin with an analysis of the results of the previous chapter, followed by consideration of factors that potentially limit the significance of our findings. We close with a discussion about future work.

5.1 Analysis

RWC Transcription Results

Comparison of our regression-based classifier results (“Regr.” columns) and our baseline classifier results (“Bayes” columns) in Table 4.3 revealed that leveraging the linear structure of log-inharmonicities yields statistically significant ($\alpha = 0.05$) performance gains over the baseline method when considering overall performance of the three RWC guitar types. Interestingly, our regression method on average outperformed the baseline only for classical guitars for all headings, while the baseline mostly outperformed the regression for only acoustic guitars guitars. Performance on just electrics was fairly evenly split, with the regression method excelling under the “Guitar-specific” heading, the baseline excelling under the “Guitar-independent” heading, and no significant difference under the “Guitar-averaged heading.”

According to our t -tests in Table 4.2, statistically significant differences at the $\alpha = 0.05$ level between the existing PCT method and our method show that for aggregated performance of classical and acoustic guitars (the “Total Overall” row), PCT retains “Guitar-specific” superiority but is overtaken by our regression method in the “Guitar-averaged” category. This suggests

that there is indeed merit to leveraging the linear structure of log-inharmonicity trajectories for improved generalized transcription. Our regression method outperforms PCT for all “Guitar-averaged” acoustic guitars, though loses slightly to PCT for classical guitars in the same scenario, perhaps implying that PCT and our regression method favor different guitar types. In general, it was rare for the error probabilities of either of our methods to not significantly differ from those of PCT.

Alternate-Tuning Results

The alternate-tuning experiments validated our tuning compensation feature, as we see in Tables 4.7 and 4.8 that the effect of appropriately scaling the inharmonicity regressions for tunings which significantly degraded standard-tuned performance is to restore transcription accuracy towards that obtained using standard tuning. Interestingly, the effect of various tunings (“DADGAD” in both electric and acoustic, “WSD” in the electric) on our guitars’ transcriptions is statistically insignificant, suggesting that we may have overestimated how problematic some alternate tunings are for inharmonicity-based systems. Also surprisingly, we learned that the “Standard” and “WSD” accuracies for our Telecaster are identical both with and without tuning compensation, and that the tuning compensation effect for “DADGAD” on our Telecaster is, though statistically insignificant, detrimental. These were perplexing results which we could not explain.

Consistency of the Regression Lines

To begin our general investigation into the factors that degraded transcription performance, we plotted every guitars’ six inharmonicity regressions side-by-side in Figure 5.1. The most prominent descriptors of these plots are the variability of the inharmonicity measurements, and the compactness or proximity of the regressions. Correlating these descriptors with the performance scores in Table 4.1 is insightful. Both acoustic guitars and classical guitars performed well, despite the more stable measurements and separated regression lines representing the acoustic guitars versus the less stable and closer locations of the regression lines representing the classical guitars. This can be explained by the fact that the strings of the classical guitars

whose regressions nearly overlap or are nearly co-linear occupy distinct pitch regions. The top left CG1 plot illustrates this best. Its two upper-left-most parallel regressions belong to nicely separated clusters of inharmonicity measurements, so classification confusion between them is unlikely because of dataset constraints that confine possible fret choices to be less than or equal to 12 and greater than or equal to 0. Hence, inharmonicity variance and regression nearness do not seem to impact transcription results greatly if the concerned regressions belong to distinguishable inharmonicity measurement clusters. For the electric guitars, however, the inharmonicity variances of the strings appear even larger, and the proximities of the regressions cause multiple overlaps (see bottom left EG1 plot). This, coupled with the fact that the concerned strings are neighbors and therefore share a higher number of plausible fretting locations, caused great difficulty for our classifier.

Figures 5.2 to 5.4 show expanded versions of some of the panels of Figure 5.1. We overlaid lines for the strings of the guitars to get a closer reading on how the regressions of each strings' vary with guitar. As can be seen in Figures 5.2, 5.3, and 5.4, the classical and acoustic guitars' inharmonicity regressions remain fairly stable from guitar to guitar, while those of the electric guitars clearly exhibit more variability, making likelihood-based classification more difficult. Electric guitar regressions belonging to strings 3 and 4 overlap considerably, which is a behavior not absent in the classical guitars (strings 1 and 4, 2 and 5, 3 and 6) and the acoustic guitars (strings 2 and 3) with some strings but which manifests in less error due to their regressions' more consistent placements than those of their electric counterparts. Indeed, the electric guitars' confusion matrices support this, and one which highlights this error trend is shown in Table 5.1.

Intrinsic Error of the Inharmonicity Measurements

We wondered to what extent our inharmonicity measurement variability was simply due to intrinsic variability in our inharmonicity measurement routine. To try to answer this, we recorded 25 uniform plucks of all strings' open notes on both our personal acoustic guitar (Takamine G-series) and a different electric guitar (Ibanez RG), and analyzed statistics of their estimated log-inharmonities. A classical guitar could not be procured at the time of the writing of this

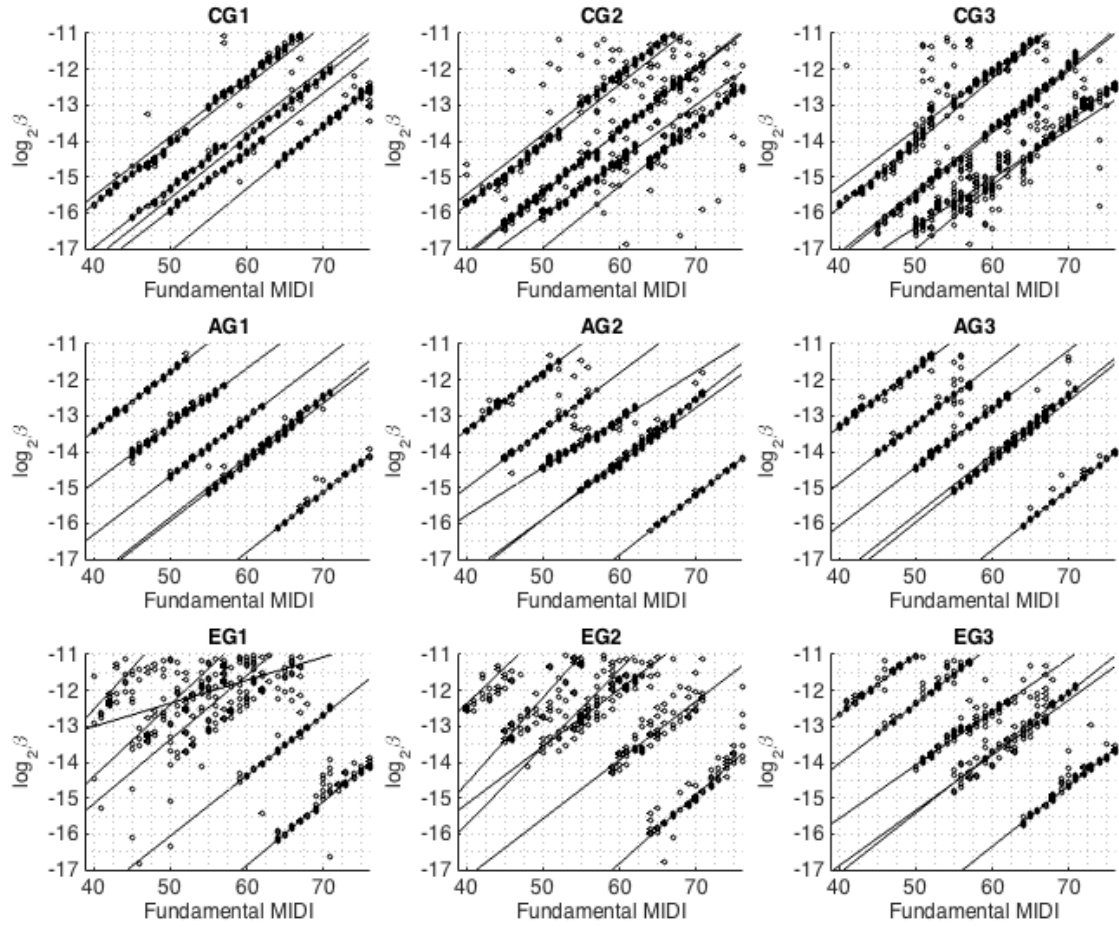


Figure 5.1: Log-inharmonicity regressions learned from the 12 recordings of each acoustic and electric guitar and the 6 selected recordings of each electric guitar superimposed over their inharmonicity measurements.

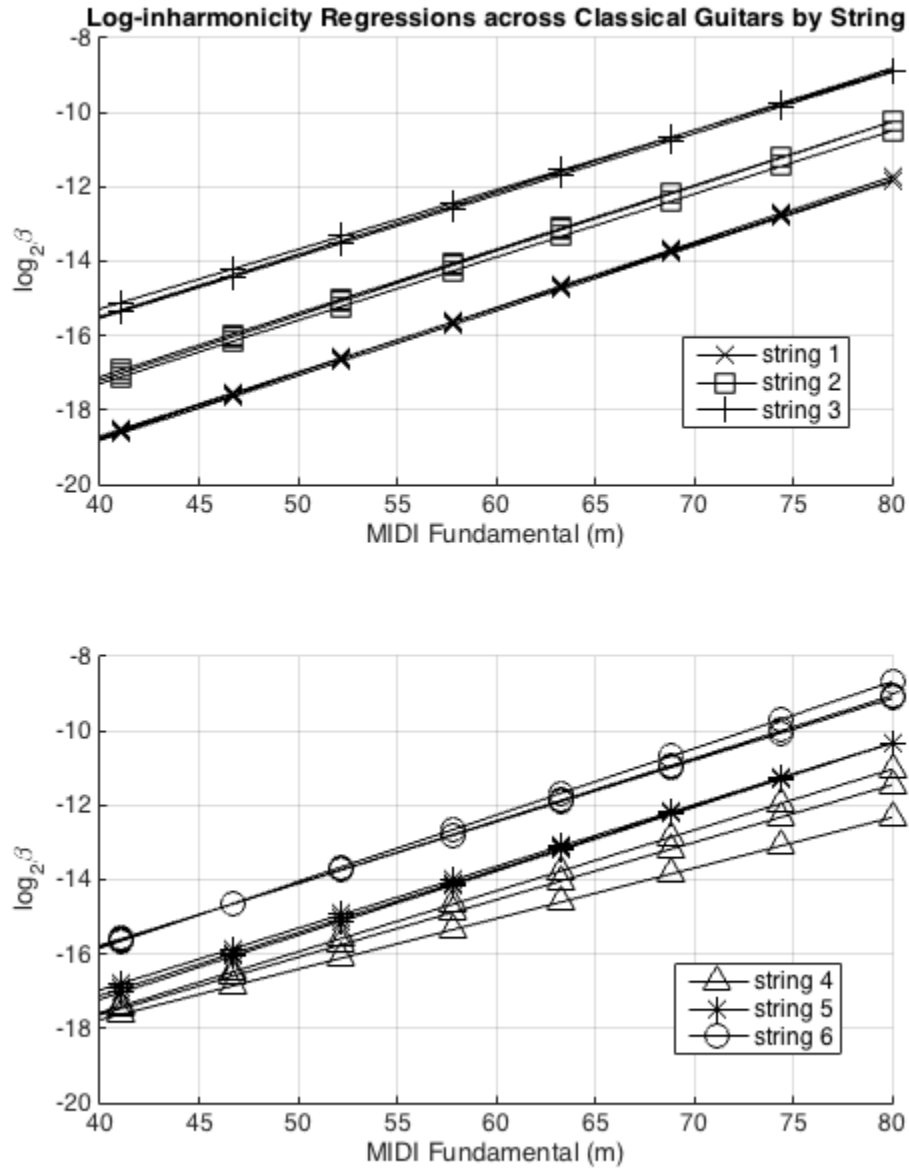


Figure 5.2: Superimposed string-wise regressions of all three RWC classical guitars. Top: Strings 1, 2, and 3. Bottom: Strings 4, 5, and 6. Plots were separated for visibility, and axes scales were kept identical for comparison. Observe the consistent intercept and slope despite varying guitar (with the only exception being String 6).

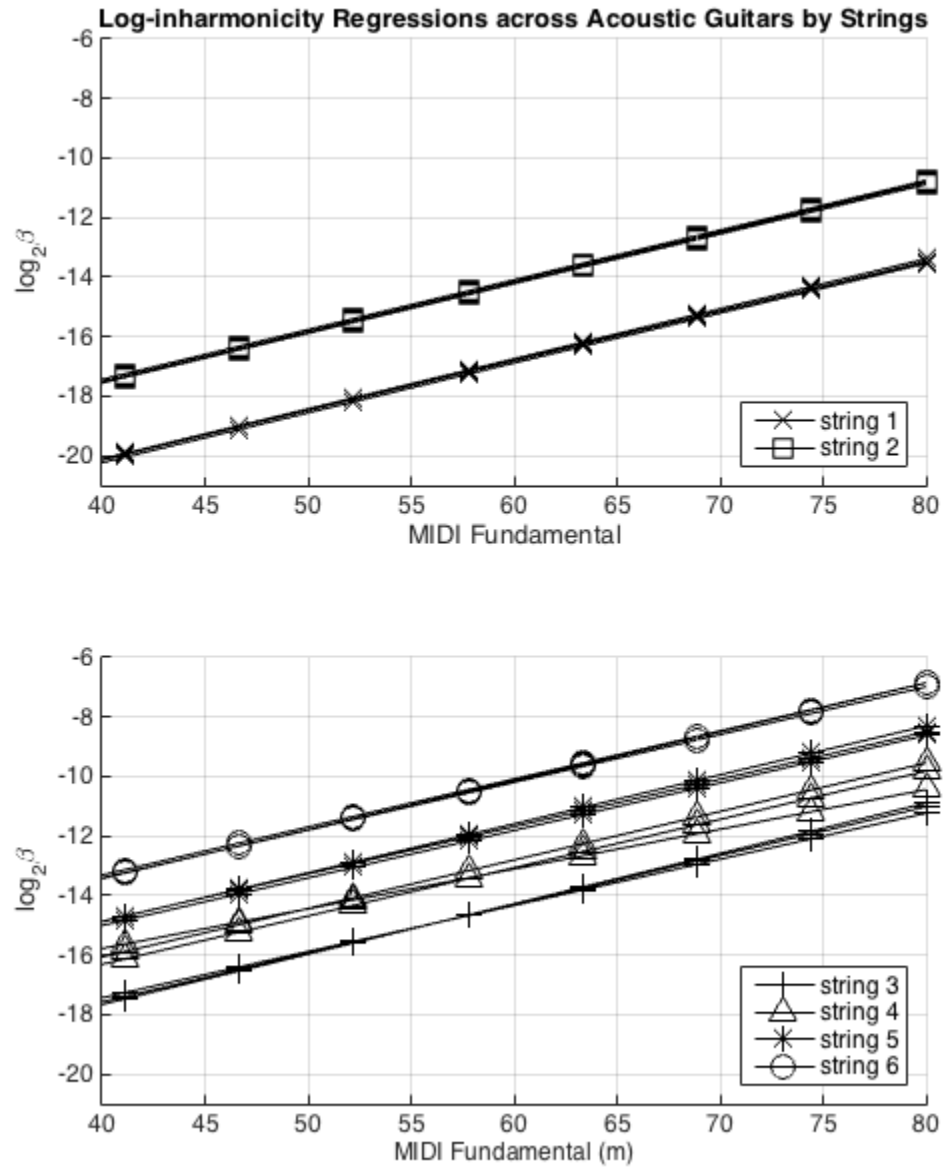


Figure 5.3: Superimposed string-wise regressions of all three RWC acoustic guitars. Top: Strings 1 and 2. Bottom: Strings 3, 4, 5, and 6. Plots were separated for visibility, and axes scales were kept identical for comparison. Observe the consistent intercept and slope despite varying guitar.

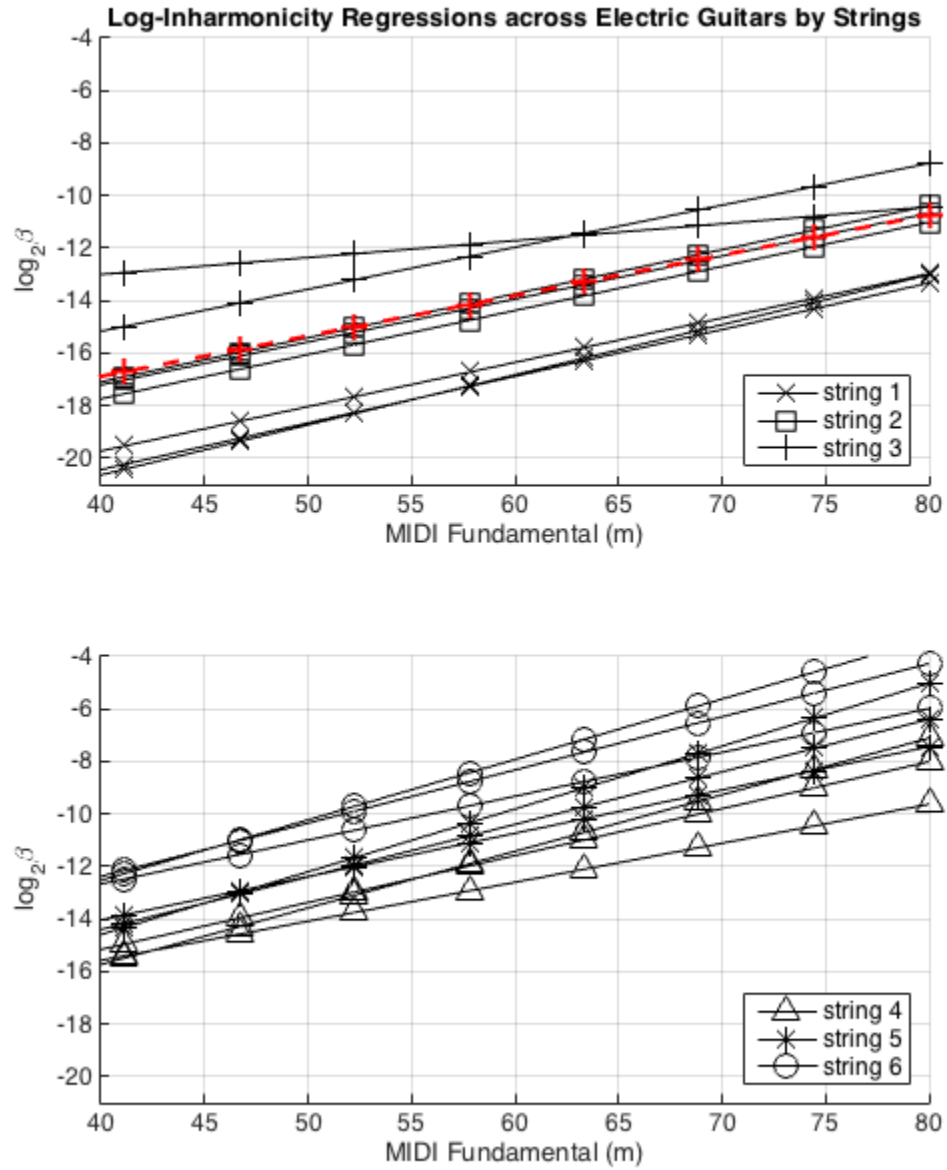


Figure 5.4: Superimposed string-wise regressions of all three RWC electric guitars. Top: Strings 1, 2, and 3. Bottom: Strings 4, 5, and 6. Plots were separated for visibility, and axes scales were kept identical for comparison. Observe the *inconsistent* intercept and slope. In particular, note the overlap between one of the regressions corresponding to String 3 (dashed red) and the regressions corresponding to String 2. Also observe the overlapping slope and intercept ranges for Strings 3 and 4 across the two plots.

	Classification					
Truth	1	2	3	4	5	6
1	1.00	0	0	0	0	0
2	0	0.77	0.23	0	0	0
3	0	0.08	0.62	0.38	0	0
4	0	0	0.23	0.46	0.31	0
5	0	0.08	0	0	0.77	0.15
6	0	0	0	0	0	1.00

Table 5.1: Confusion matrix for one classification trial with EG131. Strings 3 and 4, as they are here, were frequently misclassified because of the poor discriminability of their corresponding regressions.

section. The string gauges for the acoustic and electric were 0.012 inches and 0.010 inches, respectively. Both were played with a thick plectrum. box plots displaying the median, lower and upper quartile, data range, and any outliers of each guitar are shown in Figure 5.5 and Figure 5.6. The data ranges of the acoustic and electric guitars are on the order of one-hundredths of a unit of log-inharmonicity, which are orders of magnitude smaller than the variation we see in Figures 5.2, 5.3, and 5.4. It appears our inharmonicity measurements are fairly sound, and cannot solely account for the larger measurement variance we see, particularly in the electrics.

Nevertheless, we found that the range of determination coefficients of inharmonicity regressions for the classical and acoustic guitars was substantially smaller than that for the electric guitars. Highest and lowest quality regressions for CG091, AG111, and EG131 are shown in Figure 5.7 as an example of this; the r^2 coefficients belonging to the best and worst strings of CG091 and AG111 differ by only about one-hundredth, but the r^2 difference for EG131 is an enormous 0.87. The bottom-row of this figure suggests that, though our inharmonicity estimate is reliable, something about the RWC electric guitar EG131 is causing inconsistent measurements.

Effect of Distortion in the Pickups for Electric Guitars

We wondered if the transduction of the EG pickups somehow played a role in impeding reliable intra-guitar inharmonicity estimation – either through their transfer functions, possible noise introduction, or some non-linear characteristic. If this were the case, it would follow that

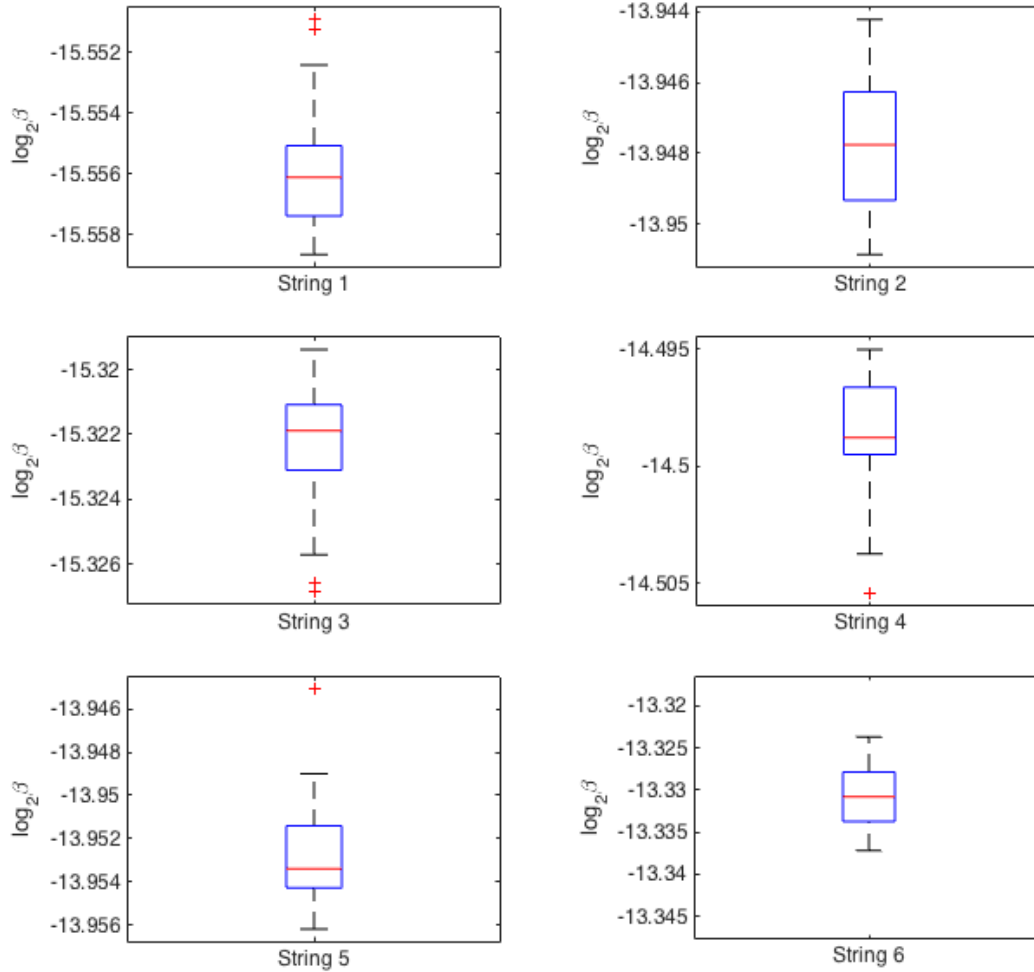


Figure 5.5: box plots of log-inharmonicity estimates for 25 uniform plucks of the six open-string notes on our acoustic guitar (Takamine G-series). Horizontal red lines are medians, horizontal blue lines are upper and lower quartiles, whiskers denote data range excluding outliers, red “+”s are outliers.

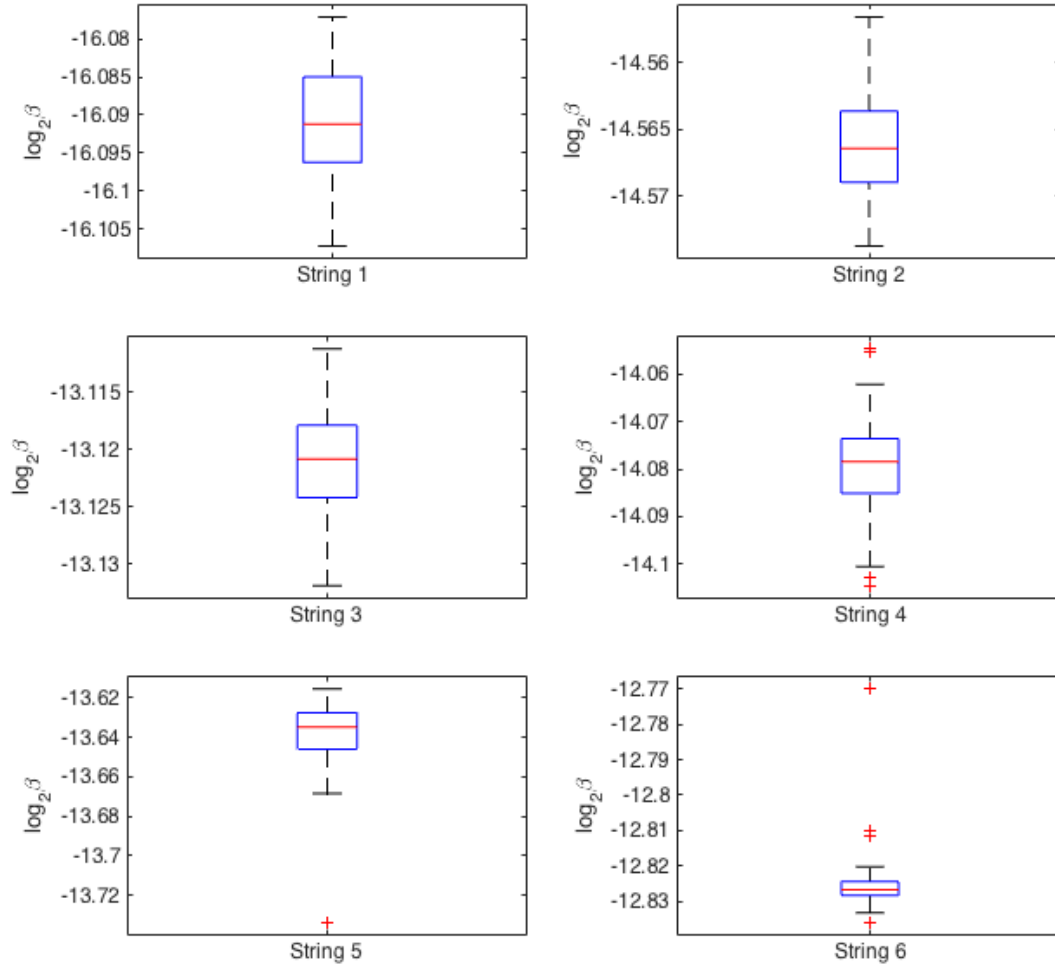


Figure 5.6: box plots of log-inharmonicity estimates for 25 uniform plucks of the six open-string notes on our electric guitar (Ibanez RG). Horizontal red lines are medians, horizontal blue lines are upper and lower quartiles, whiskers denote data range excluding outliers, red “+”s are outliers.

Best-Worst Fit Comparisons for Log-Inharmonicity Regressions

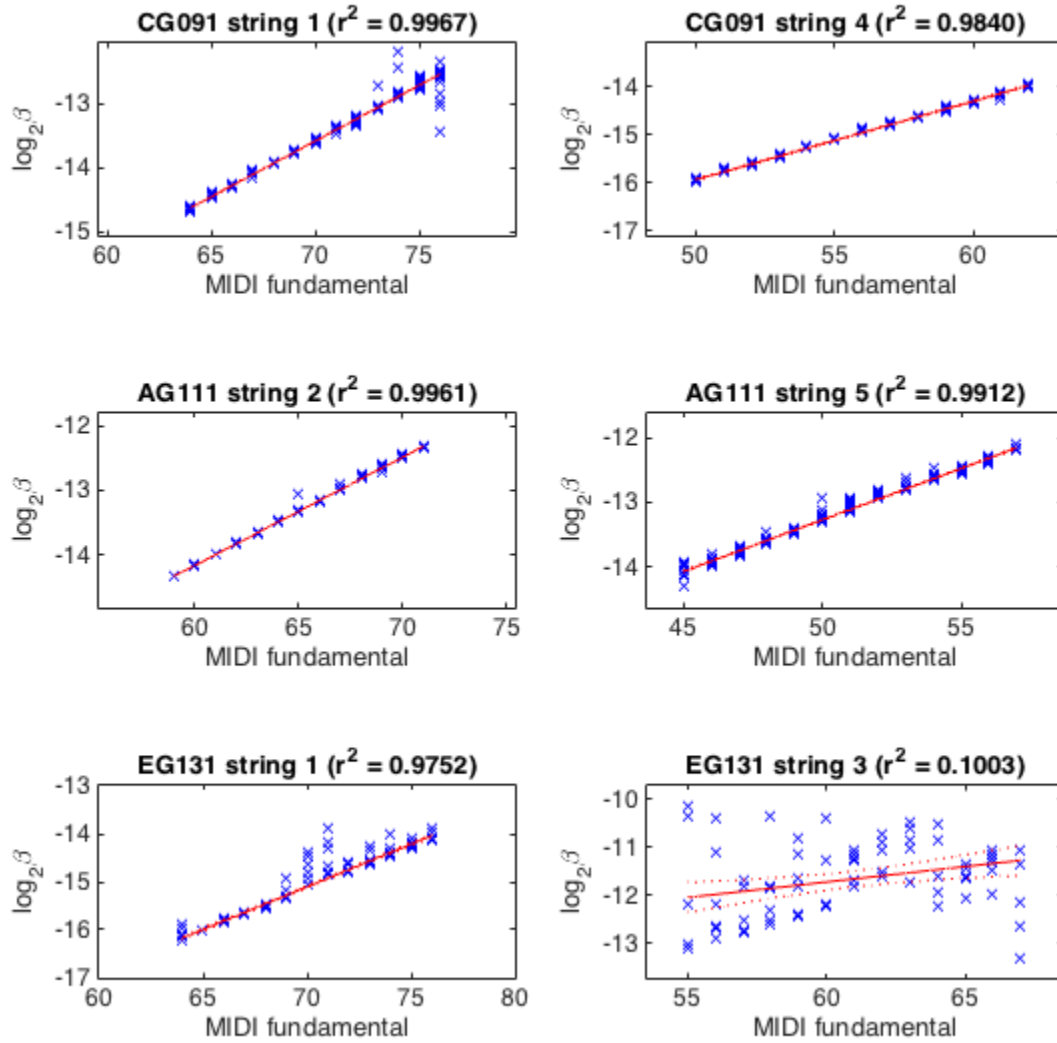


Figure 5.7: Best and worst bisquare log-inharmonicity regression fits, as measured by r^2 , for representative guitars from each of the three classes. Top row: classical. Middle row: acoustic. Bottom row: electric. Left and right columns are best and worst fits, respectively. Dashed red lines, only visible on the bottom-right plot, denote 95% confidence intervals.

higher-intensity performances would elicit this impeding phenomenon to a greater extent and produce measurably more variable inharmonicity estimates than would lower-intensity performances. Since one of the RWC musicality parameters that varies for each featured guitar is dynamic level of the performance, we were able to separate higher- and lower-intensity performances of the same guitar and analyze this interaction between degree of pickup transduction and inharmonicity estimation variability. For each guitar, and for each string, we collected note performances by common dynamic levels (piano, mezzo, and forte) and regressed their log-inharmonicity against their fundamentals, as usual. We then plotted statistics of the residuals of the log-inharmonicity against their played performance intensity. The residual of a datapoint is the error between its value and the approximation of its value by its representative regression. Figure 5.8 illustrates these statistics for EG131; other electric guitars' box plots were omitted because their statistics were similar. We expected that residual variability would increase with dynamics, but the data didn't support this hypothesis. Interquartile ranges didn't exhibit any clear trend with respect to performance intensity, nor was the magnitude of their variation noteworthy.

Interestingly, neither the reliability of our inharmonicity estimation nor the dynamic level of the performance explain the large variance observed in inharmonicity estimates in the one electric guitar in Figure 5.7. Plucking with the finger versus with a plectrum was one more variable present in RWC recordings of a single guitar, but wasn't investigated in this work, so it remains unclear why such large variance persists in measurements on a single guitar. Possible other factors are numerous: fret buzz, plectrum thickness, variability in the gripping strength of the fret-depressing hand, etc.

Effect of Guitar-String Geometry on Inharmonicity Estimates

Many more factors could also be contributing to the inter-guitar variance of the inharmonicity measurements. Differences in the physical properties of the strings of various RWC guitars (like gauge, whether the string is wound or unwound, length due to intonation adjustments, stiffness due to age) could be vast, and each of these factors are indeed components in the definition of inharmonicity in equation (2.2). Details regarding these instrument-specific factors aren't

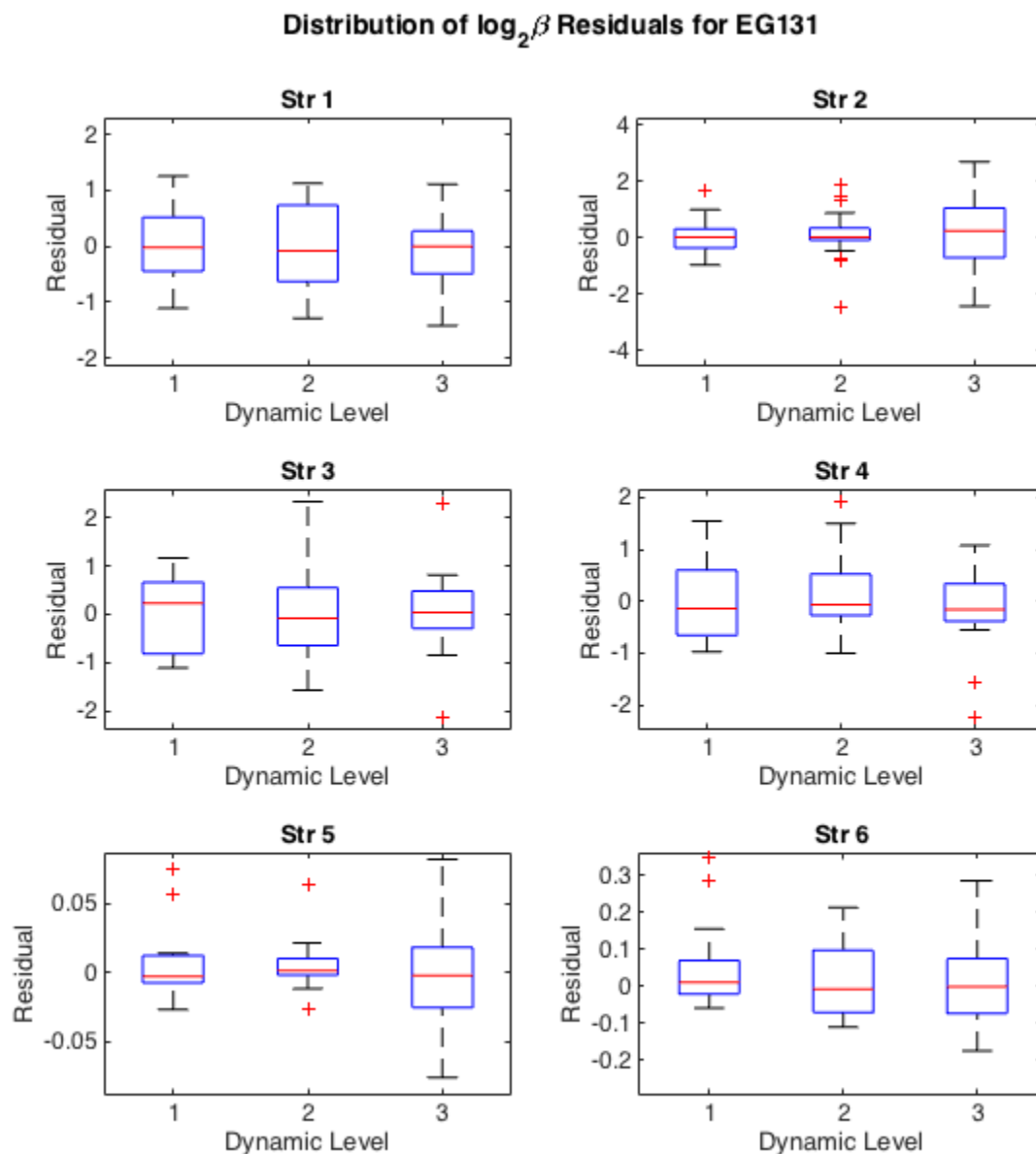


Figure 5.8: box plots of log-inharmonicity regression residuals for EG131 by dynamic level. Group 1: piano, Group 2: mezzo, Group 3: forte. Red center horizontal lines are medians; top and bottom edges of blue boxes represent upper and lower quartiles; whiskers depict full range of data excluding outliers; red plus signs denote outliers. Vertical axes not constant throughout plots; we're highlighting relative distributions of the box plots here.

recorded in the RWC annotations, so it is plausible that the electric guitar transcription difficulty of our system is exacerbated by unknown combinations of these confounding factors.

To get an idea of the magnitude of inharmonicity changes imparted by different values of these factors, we did rough calculations and took some measurements of our guitar to confirm them. For string gauge d , a reasonable spread of values one might encounter on electric guitars is from a light 0.008 inches to a heavy 0.012 inches. These numbers describe the diameter of the first string, though, by convention, they also indirectly refer to the relative thicknesses of the remaining five strings. The percent increase in the diameter of the first string from lightest to heaviest gauge is approximately 67%. As an aside, it is important we consider a string whose winding remains constant – either wound or unwound – throughout gauge sizes, as winding will presumably affect string stiffness as well. Referring to the inharmonicity definition in equation (2.2), we see that the inharmonicity is proportional to the fourth power of diameter,

$$\beta \propto d^4. \quad (5.1)$$

If we insert a scaling factor a on the string diameter, we obtain

$$\beta \propto (ad)^4 \quad (5.2)$$

$$\Delta\beta \propto a^4 \quad (5.3)$$

$$\Delta \log_2 \beta = 4 \log_2 a \quad (5.4)$$

showing us that, for a maximum string diameter increase of $a = 1.67$, we can expect a variation in log-inharmonicity of approximately $4 \log_2 1.67 \approx 3$ units.

For string length l , maximum extension or contraction length for intonation adjustment purposes probably falls within 10mm. We know from equation (3.16) that the fundamental of a

string is proportional to the square root of its tension and inversely proportional to its length,

$$f_0 \propto \frac{\sqrt{T}}{L}, \quad (5.5)$$

so a string whose length changes by a factor of b would have to experience an increase in tension of b^2 to maintain the same open pitch. Again, referring to the inharmonicity definition in equation (2.2), we see that inharmonicity is inversely proportional to tension and the square of length, or

$$\beta \propto \frac{1}{Tl^2}. \quad (5.6)$$

Propagating the scaling factor b into log-inharmonicity estimation gives us

$$\beta \propto \frac{1}{Tl^2b^4} \quad (5.7)$$

$$\Delta\beta \propto \frac{1}{b^4} \quad (5.8)$$

$$\Delta \log_2 \beta = -4 \log_2 b. \quad (5.9)$$

In words, for a scaling factor b on string length, we can expect our log-inharmonicity measurements to shift by $-4 \log_2 b$. For a 10-mm intonation extension on a string which measures a conservative 600 mm, b amounts to roughly 1.0167 yielding inharmonicity perturbations of $-4 \log_2 b \approx -0.0956$. Rounding up, we would expect maximum log-inharmonicity shifts of one-tenth due to intonation length adjustments.

String stiffness Q is harder to predict, since it presumably depends on multiple factors, like whether the string is wound and its age. For this reason, we don't obtain approximate expectations of log-inharmonicity shifts for this factor. We proceed with measurements of these various parameters to check the validity of our expectations.

For our measurements, we recorded 0.009-in, 0.010-in, and 0.011-in gauge sets of the unwound third string of a personal electric guitar (Ibanez RG), as well as another take of a 0.011-in

set with a nickel-wound third string. We also had two sets of 0.010-in set gauges, one of which was old and the other which was new, in an attempt to examine the effect of possible age-related stiffness. Finally, we recorded each of the previous settings at two lengths: the owner’s default 64-cm intonation length and a 0.5-cm intonation extension. box plots displaying the variation in inharmonicity measurements due to these factors are shown in Figure 5.9.

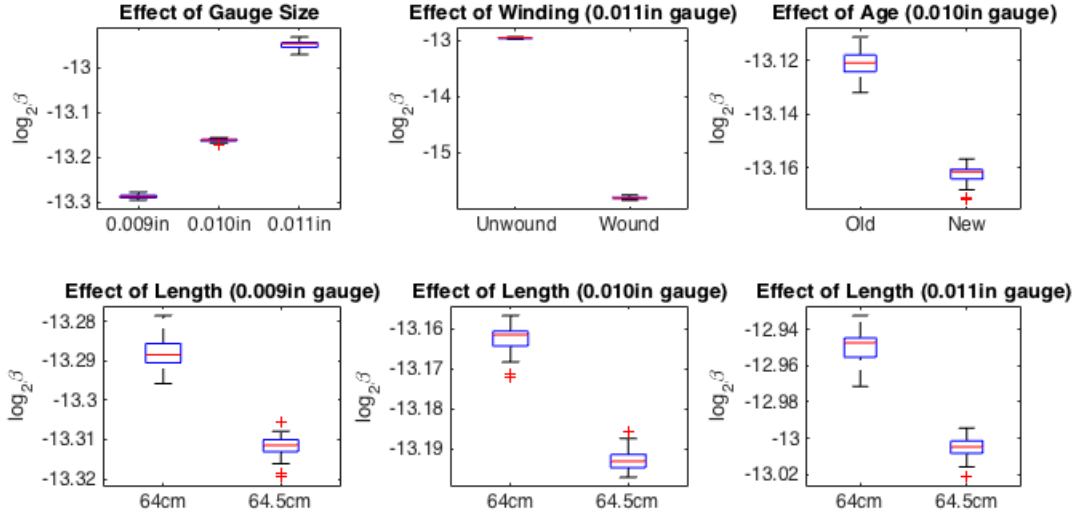


Figure 5.9: Effects on log-inharmonicity measurements of various electric guitar geometry parameters. In all box plots, measurements are from 25 uniformly-played open notes (G3) on string 3 in standard tuning. Horizontal red lines are medians, horizontal blue lines are upper and lower quartiles, whiskers denote data range excluding outliers, red “+”s are outliers. Note the changes in scale on the vertical axes.

The plots in Figure 5.9 for gauge size and length agree with our expectations. The shifts in log-inharmonicity between our measured 0.009-in and 0.011-in gauges fall within the 3-unit predicted spread, and the shifts in our 5-mm length adjustments also fall within our rough estimations. Length and age-related stiffness impart log-inharmonicity changes on the order of only hundredths or tenths, which do not seem large enough to perturb the regressions in Figure 5.4 such that significant increases in transcription error occur; the stable regressions (belonging to strings 1 and 6, for example) are separated from their neighboring regressions by about 2 units, so it seems unlikely that inharmonicity biases due to age-related string stiffness or intonation-

motivated length differences would compromise the performance of our system. Nevertheless, the effect of string winding (top row, middle plot of Figure 5.9) is substantial: log-inharmonicity measurements for a nickel-wound third string were about 3 whole units less than its unwound counterpart. It therefore seems clear that extreme gauge size differences and the presence of winding have the potential to considerably alter the placement of regressions of the strings and compromise their predictive performance. This could also shed light on the outsized variance exhibited by the regressions corresponding to the third electric guitar string in the top plot of Figure 5.4. The magnitude of the deviations of the third string across guitars is commensurate with that of winding- and gauge-related inharmonicity changes, so it is plausible that this is the case in some of the RWC electric guitars. Overall, these plots tell us that the different guitar geometry factors – string gauge, winding, stiffness, and intonation length – indeed considerably affect our inharmonicity measurements.

Effect of Overdrive on Inharmonicity Estimates

Lastly, we examined the consequences of an overdrive effect on inharmonicity estimation. Though not immediately concerned with this work (the RWC dataset we used is entirely clean), it is a relevant question that seeks further insight into inharmonicity-based transcription systems’ practicality in real-world guitar recordings. Briefly, overdrive is a phenomenon that occurs when guitarists play at gain levels that saturate the vacuum tubes in their amplifiers. The tubes non-linearly distort the guitar signal, producing harmonics and overtones that lend a warm and increasingly “broken-up” tone to their instrument. This analog phenomenon is an iconic facet of the sound of the electric guitar, and software plug-ins and pedals that attempt to simulate the effect without vacuum tubes are abundant. We applied two levels (12 dB and 24 dB) of Logic Pro’s stock Overdrive plug-in to the 25 open-note recordings of String 3 of our Ibanez RG. The tone knob was fixed to 980 Hz and the output level was set to 0 dB. For each trial, we measured log-inharmonities as before and plotted their distributions in Figure 5.10.

We see that the effect of Logic Pro’s overdrive plug-in on log-inharmonicity is to shift the median estimates upward by a few tenths and to increase the data range by an order of magni-

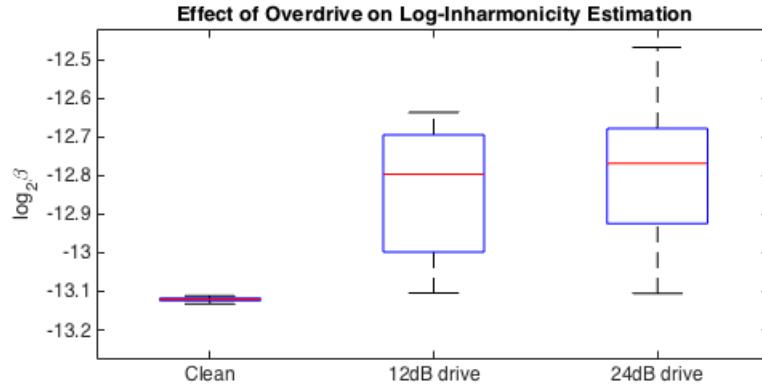


Figure 5.10: box plots of log-inharmonicity measurements of 25 open-note plucks on a new, 0.010-in gauge, unwound third string, processed with two levels of drive in Logic Pro’s stock overdrive plug-in.

tude. Interestingly, the 24-dB trial produces a more focused interquartile range of measurements, though the data range increases further still. Not shown beyond the bounds of this plot were two outliers for each of the 12-dB and 24-dB trials – these were omitted for greater box plot visibility. It appears that the relative degradation of inharmonicity measurement quality with the introduction of a digital overdrive is appreciable, though we’re unsure of the degree to which this might affect inharmonicity-based transcription results in practice.

5.2 Limitations

It is important to highlight factors that potentially limit the significance of our findings. Our system currently operates only on clean (i.e. without effects processing), isolated, monophonic guitar recordings, which are obviously not representative of the majority of guitar audio. Multiple layers of effects processing are commonly used by performers of the instrument to achieve distinctive, creative sounds. Guitarists often play in concert with other musicians, producing complex soundscapes instead of being featured in isolation. Moreover, guitar passages are rarely in their entirety strictly monophonic. These are large obstacles that separate our RWC subset from actual guitar performances, and thus the implications of our transcription results should be

considerably tempered. Noteworthy, though, is the extent to which other teams have overcome some of these factors. Certain tablature transcription systems [? ? ? ?] have implemented polyphonic capabilities into their systems with good success, so polyphony is perhaps a smaller barrier than the others. As part of a multipitch estimation system in [?], Yazawa deduced tablature from RWC recordings that featured guitar passages playing over backing tracks, though the tablature accuracy is not reported.

Additionally, we assumed perfect fundamental pitch estimation in this work. Our decision to simply use pitch labels of the RWC recordings stemmed from convenience and scope; the focus of this work was simply evaluation of a specific approach to examining notes' inharmonicities for tablature transcription, so not concerning ourselves with the reliability of a pitch detection routine was a luxury that expedited this thesis. That being said, it should be acknowledged that the reliability of one's pitch estimator is a sure factor in the translation of the results of this work to the real world.

Furthermore, the inharmonicity estimation routine we used, MAT, was different from that used by Barbancho. The credibility of the comparison between our results and theirs is thus somewhat compromised. It is quite possible that any performance gains observed in our implementation are due in part to potentially superior inharmonicity extraction. This is, unfortunately, a factor we did not control for, and if time had allowed we would have replicated Barbancho's system and substituted their inharmonicity estimation step with MAT to produce results that were as comparable as possible.

5.3 Future Work

Inharmonicity-based tablature is a fascinating topic, and future researchers should advance some of the findings in this thesis. In the previous section, we confirmed that inter-guitar string geometry parameters – gauge size, length, winding, age – played a significant role in inter-guitar inharmonicity measurement variability. This makes clear that understanding and accounting for

the factors that perturb the inharmonicity measurements of a system is key for good performance, and methods to exploit these deterministic perturbations should be developed. Perhaps learning separate inharmonicity trends for various commonly-encountered combinations of these factors would make inharmonicity-based systems more viable. There's also much we did not solve about our intra-guitar inharmonicity variability, and answers should be sought to why inharmonicity variability remains high on just one RWC guitar in some cases.

We also need to more deeply understand how inharmonicity measurements are affected by the intentional modulation of guitar audio common in real-world performances. Creative effects processing with distortion, reverb and delay, for example, are a ubiquitous feature of the guitar's performance, and inharmonicity-based systems need to show that reliable measurement amidst this processing is attainable if they are to claim any practical merit. Thorough characterization of the consequences of string-bending, neck-bending, palm-muting, vibrato, and other performance variations is also needed, since these too are common features of guitar performances that probably obscure the inharmonicity estimation process.

Future work should also consider ways to improve performance on the system level. One potential area for run-time performance enhancement is the use of musical context. Musical passages on the guitar tend to exhibit continuity; it is usually the case that a riff is played in a localized area of the fretboard, for example, instead of jumping around the length of the guitar neck. Transcription systems could apply these performance-borne simplifications to their output to potentially improve accuracy, say, when a stray note is transcribed a plausible yet questionable twelve frets away from the rest of those in the riff. A measure of objective confidence in the transcription is required to do this. Without it, there'd be no way to determine whether it was the stray note or the majority of the riff which was in fact correctly transcribed. To this end, we investigated the amenability of our regression classification scheme to reliable confidence measurements, and found encouraging results. Figure 5.11 displays the probability, for each of the nine RWC guitars, of our system producing a correct fretboard transcription given increasingly wide windows in which we consider the inharmonicity estimate of a note to have

fallen. The inverse trend between residual window width and correctness probability shows that we can reliably infer some degree of confidence in a note's transcription given its distance from its classified line. This should be leveraged in a post-processing stage in later iterations of this and others' work to further refine tablature with isolated erroneous, yet plausible, transcriptions.

Many extensions specific to this thesis can be explored in further work as well. Considering the good performance of the baseline Bayes classifier, experimentation should be continued with other standard classifiers. Perhaps a different well-studied classification routine that we've overlooked is even better-suited to the structure of the inharmonicity data, and would outperform our regression-based likelihood approach. Additionally, varying the regression polynomial could be an interesting experiment. Perhaps the system would perform differently if the log inharmonicity transformation was forgone and an exponential fit was explored instead. Exploration of different signal-processing parameters for inharmonicity extraction and frame aggregation, for example window choice and hop size, could be useful. Fitting our regression system with polyphonic capability and observing the extent to which performance changes would also be a worthwhile study.

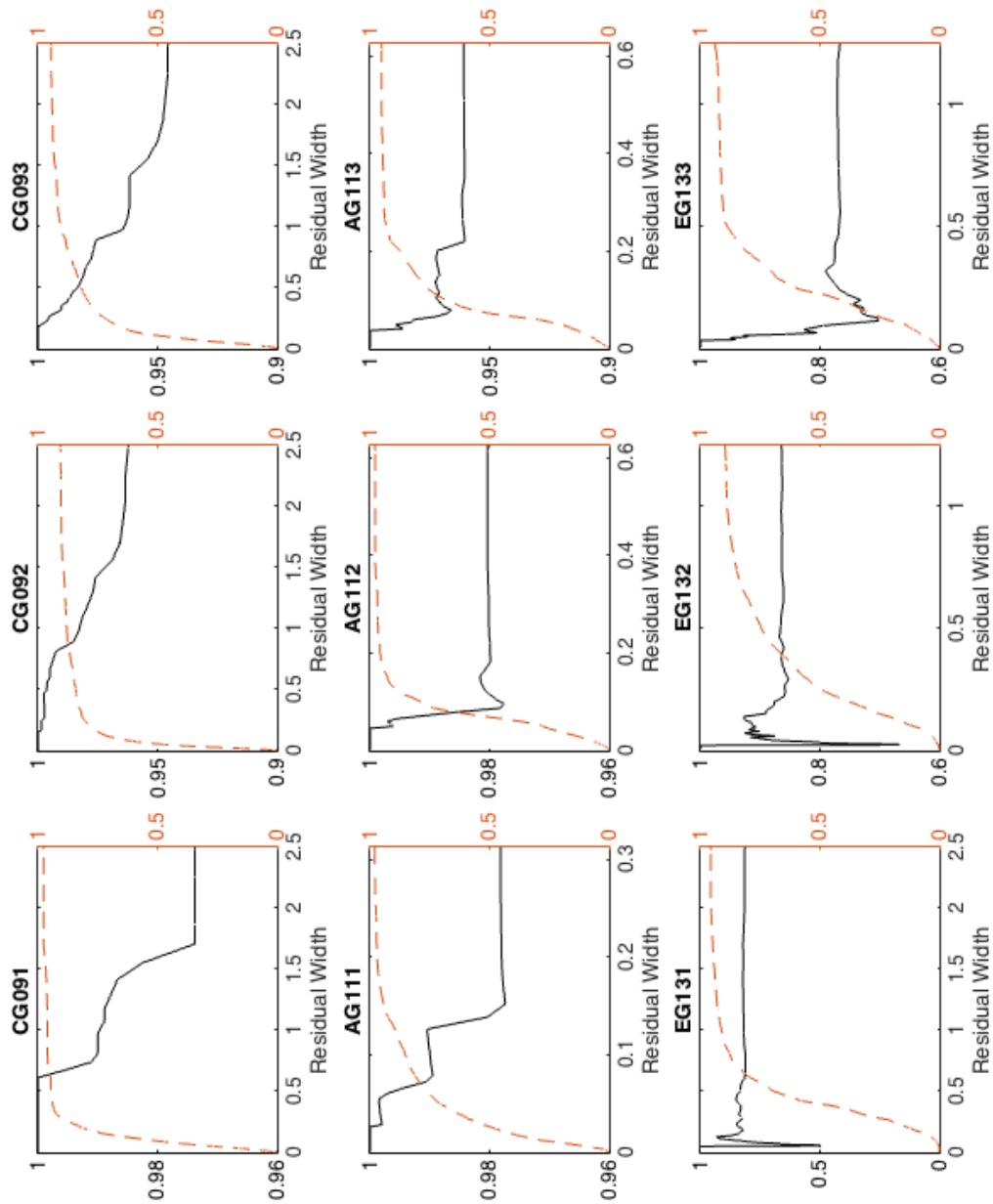


Figure 5.11: Solid black lines and left vertical axis: probability of correctly transcribing a note given the residual width under which the log-inharmonicity of the note falls. Dashed orange lines and right vertical axis: CDF of the log-inharmonicity residuals, so one can get a sense of the proportion of data represented by the horizontal axis.

Chapter 6: Summary

We began by introducing the guitar and an essential component of its musicality: the overlapping pitch ranges of its strings. We related this to tablature and its ability to uniquely specify fretboard positions for musical passages, then highlighted its popularity and its tedious manual annotation process. This led us to introduce automatic transcription systems and our goals for this work as an extension to the topic, which we reviewed in the following chapter.

We paid special attention to the partial coincidence tally (PCT) transcription method introduced by Barbancho in [?], which exploited only one feature known as inharmonicity in development of a successful system using only audio. We explained how inharmonicity affects the degree of upward skew of a note’s partials, then discussed the intuition behind its discriminative power before introducing our novel approaches.

After reiterating one of Barbancho’s key realizations – that inharmonicity along a given string was deterministic – we showed that the logarithm of these trajectories was linear with respect to MIDI pitch. We consequently proposed characterizing guitar strings as their log-inharmonicity trajectories and performing classification (and therefore implicitly tablature transcription) by probabilistically assigning to unknown notes the strings whose lines best approximated the notes’ log-inharmonicity. Following this, we proposed a baseline Bayesian method against which we could evaluate the utility of our regression innovation, and which similarly operated by assigning to an unknown note the fretboard position which maximized the probability of having observed its measured inharmonicity. Lastly, we derived the effect of a tuning change on our regression method, and arrived at a general tuning-compensation feature applicable to any inharmonicity-

based system.

Next, we reported results from experiments on a subset of the RWC music instruments database and a couple personal guitar recordings. Comparisons of our baseline and regression approaches against the PCT method revealed that, overall, the regression method achieves lower error rates when transcribing guitars on which it wasn't individually trained. It also outperformed our baseline, arguing that leveraging inharmonicity trajectories is a worthwhile pursuit. We also validated our inharmonicity compensation feature, reporting accuracy increases on personal recordings of alternately-tuned electric and acoustic guitars.

A cursory investigation of the cause for the poor electric guitar transcription performance revealed that inharmonicity estimation for the instruments is less reliable, and both the pickups of the electric guitars and the intrinsic variability of the measurement routine are not to blame. We briefly examined the effect of various factors, such as guitar string geometry (gauge, length, stiffness, winding) and effects processing (overdrive), on inharmonicity measurement reliability and found that their influences were considerable. This potentially helps to explain the high inter-guitar inharmonicity variance, but it is still unclear why intra-guitar inharmonicity variance remains high for some of the RWC guitars' strings. We discussed limitations of this work, cautioning against misinformed application of our reported results to real-world guitar scenarios, and closed with future work, encouraging development of systems which further study and exploit both guitar string geometry and effects processing factors, as well as potential system-level improvements and additions.

In conclusion, characterizing guitar strings as log-inharmonicity trajectories against pitch and classifying notes by maximizing the observation probabilities of the trajectories is an effective approach to inharmonicity-based tablature transcription. Most notably, in addition to outperforming the benchmark probabilistic classifier introduced in this work, it overall surpasses the existing inharmonicity-based method for classical and acoustic guitars on which the system wasn't exclusively trained. Furthermore, adjusting inharmonicity estimates with the compensation factor introduced in this work is effective for improving transcription accuracy on alternately-tuned

guitars.

